STIC-Biotech/ChemLib 170169 mg

From: Marvich, Maria
Sent: Tuesday, November 01, 2005 6:33 AM
To: STIC-Biotech/ChemLib
Subject: 09/463542

RECEIVED
NOV - 1 2005
TECHNOLOGY CENTER 1600/2900
(STIC)

Please search including interference 1-125 of SEQ ID NO:1
818-1320 of SEQ ID NO:3

368--1144 of SEQ ID NO:34

Thank you

Maria Bonovich Marvich, PhD
United States Patent and Trademark Office
Remsen 2B84
AU 1636
Mail Box 2C70
571-272-0774

1633?

****
Searcher:_____
Searcher Phone:_____
Date Searcher Picked up:_____
Date completed:_____
Searcher Prep Time:_____
Online Time:_____

****
Type of Search
NA#_____ AA#:_____
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure #:_____Text:_____
Inventor:_____ Litigation:_____

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other (Specify):_____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:52:53 ; Search time 2639.94 Seconds
(without alignments)
9232.381 Million cell updates/sec

Title: US-09-463-542-3_COPY_818_1320
Perfect score: 503
Sequence: 1 tagagtaagtaccttaggaa.....atatttggaaactgatgtct 503

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	473	94.0	597	9	AF012874	AF012874 Homo sapi
	2	455.2	90.5	984	9	AF310249	AF310249 Homo sapi
	3	455.2	90.5	1100	6	AR121474	AR121474 Sequence
	4	455.2	90.5	1100	9	AB005520	AB005520 Homo sapi
c	5	455.2	90.5	85873	6	AX951686	AX951686 Sequence
	6	455.2	90.5	135675	9	AC093174	AC093174 Homo sapi
	7	455.2	90.5	148828	9	AY157024	AY157024 Homo sapi
c	8	455.2	90.5	166049	9	AC090947	AC090947 Homo sapi
c	9	362.8	72.1	186028	2	AC151849	AC151849 Callithri
c	10	219.6	43.7	191540	2	AC136055	AC136055 Rattus no
c	11	219.6	43.7	245724	2	AC091418	AC091418 Rattus no
c	12	219.6	43.7	250169	2	AC120668	AC120668 Rattus no
	13	216.4	43.0	682	10	S79407	S79407 mPPAR gamma
	14	216.4	43.0	1159	10	AY243584	AY243584 Mus muscu
	15	216.4	43.0	1160	10	AY236531	AY236531 Mus muscu
	16	216.4	43.0	185737	2	AC125447	AC125447 Mus muscu
c	17	216.4	43.0	237588	2	AC025581	AC025581 Mus muscu
	18	216	42.9	1152	10	AY236530	AY236530 Mus muscu
	19	212.2	42.2	2982	10	AY389536	AY389536 Mus muscu
	20	212.2	42.2	2983	10	AY389535	AY389535 Mus muscu
	21	202.8	40.3	3003	10	AY389534	AY389534 Mus muscu
	22	165	32.8	212216	2	AC131898	AC131898 Oryctolag
	23	155.6	30.9	634	4	AY044238	AY044238 Sus scrof
	24	147	29.2	1077	4	AF288373	AF288373 Bos tauru
	25	96.6	19.2	218	4	SSC6758	AJ006758 Sus scrof
	26	95	18.9	154195	2	AC142242	AC142242 Atelerix
	27	47.2	9.4	160766	5	BX088554	BX088554 Zebrafish
c	28	45.2	9.0	176854	2	AC119375	AC119375 Rattus no
	29	45.2	9.0	200000	9	AB019438	AB019438 Homo sapi
c	30	45	8.9	215362	10	AL663108	AL663108 Mouse DNA
	31	44.4	8.8	165260	9	AC024341	AC024341 Homo sapi
c	32	44.2	8.8	3162	6	AX430407	AX430407 Sequence
	33	44.2	8.8	166016	9	AL591378	AL591378 Human DNA
	34	44.2	8.8	174649	2	AC068210	AC068210 Homo sapi
c	35	44	8.7	143618	2	AC026855	AC026855 Homo sapi
	36	44	8.7	181004	2	BX664751	BX664751 Danio rer
	37	44	8.7	195319	2	CR388410	CR388410 Danio rer
	38	43.8	8.7	47328	3	AF295546	AF295546 Malawimon
	39	43.8	8.7	51700	9	AL590243	AL590243 Human DNA
c	40	43.8	8.7	74427	9	AL358796	AL358796 Human DNA
c	41	43.8	8.7	180619	9	AL356513	AL356513 Human DNA
c	42	43.8	8.7	181632	2	AC053508	AC053508 Homo sapi
	43	43.8	8.7	183118	2	AL355528	AL355528 Homo sapi
c	44	43.6	8.7	6215	6	AX346093	AX346093 Sequence
c	45	43.4	8.6	105593	8	AC144483	AC144483 Medicago

ALIGNMENTS

RESULT 1
AF012874
LOCUS AF012874 597 bp DNA linear PRI 03-MAY-1999
DEFINITION Homo sapiens PPARG gene, isoform 2, promoter and 5'UTR sequence.
ACCESSION AF012874
VERSION AF012874.1 GI:4731428
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 597)
AUTHORS Fajas,L., Auboeuf,D., Raspe,E., Schoonjans,K., Lefebvre,A.-M.,
Saladin,R., Najib,J., Laville,M., Fruchart,J.-C., Deeb,S.,
Puig-Vidal,A., Flier,J., Briggs,M., Vidal,H., Staels,B. and
Auwerx,J.
TITLE The organization, promoter analysis, and expression of the human
PPARgamma gene
JOURNAL J. Biol. Chem. 272 (30), 18779-18789 (1997)
MEDLINE 97373577
PUBMED 9228052
REFERENCE 2 (bases 1 to 597)
AUTHORS Fajas,L., Auwerx,J., Saladin,R. and Briggs,M.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1997) Cardiovascular Res., Ligand
Pharmaceuticals, 9393 Towne Centre Dr., San Diego, CA 92121-3016,
USA
FEATURES Location/Qualifiers
source 1. .597
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
gene 1. .>597
/gene="PPARG"
/note="peroxisome proliferative activated receptor, gamma"
promoter 1. .463
/gene="PPARG"
5'UTR 464. .597
/gene="PPARG"
/note="isoform 2"
ORIGIN

```
  Query Match           94.0%;  Score 473;  DB 9;  Length 597;
  Best Local Similarity  100.0%;  Pred. No. 5e-101;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 CAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTC 90
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTC 60

Qy         91 TTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTT 150
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTT 120

Qy        151 TGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAG 210
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAG 180

Qy        211 CCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGA 270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGA 240

Qy        271 TAATTCTAAATACAGTACAGTTCACGCCCCTCACGAGACACTGAACATGTGGTCACCGGC 330
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAATTCTAAATACAGTACAGTTCACGCCCCTCACGAGACACTGAACATGTGGTCACCGGC 300

Qy        331 GAGACAGTGTGGCAATATTATCCCTGTAATGTACCAAGTCTTGCCAGAGCAGTGAACATT 390
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAGACAGTGTGGCAATATTATCCCTGTAATGTACCAAGTCTTGCCAGAGCAGTGAACATT 360

Qy        391 ATGACACAACTTTTTGTCACAGCTGGCTCCTAATAGGACAGTGCCAGCCAATTCAAGCCC 450
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATGACACAACTTTTTGTCACAGCTGGCTCCTAATAGGACAGTGCCAGCCAATTCAAGCCC 420

Qy        451 AGTCCTTTCTGTGTTTATTCCCATCTCTCCCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTCCTTTCTGTGTTTATTCCCATCTCTCCCAAATATTTGGAAACTGATGTCT 473


RESULT 2
AF310249
LOCUS       AF310249                984 bp    DNA     linear   PRI 12-MAR-2001
DEFINITION  Homo sapiens peroxisome proliferator activated receptor gamma 2
            gene, upstream sequence.
ACCESSION   AF310249
VERSION     AF310249.1  GI:13274397
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 984)
  AUTHORS   Lim,C.H., Kok,C.C., Samian,M.R., Najimudin,N. and Tengku
            Muhammad,T.S.
  TITLE     Molecular cloning and sequencing of the human Peroxisome
            Proliferator Activated Receptor Gamma 2 promoter
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 984)
  AUTHORS   Lim,C.H., Kok,C.C., Samian,M.R., Najimudin,N. and Tengku
            Muhammad,T.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2000) School of Biological Sciences, Universiti
            Sains Malaysia, Minden, Penang 11800, Malaysia
FEATURES             Location/Qualifiers
     source          1. .984
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     misc_feature    1. .982
                     /note="contains promoter and 5' untranslated region of
                     peroxisome proliferator activated receptor gamma 2 gene"
     mRNA            <982. .>984
                     /product="peroxisome proliferator activated receptor gamma
                     2"
     CDS             982. .>984
                     /codon_start=1
                     /product="peroxisome proliferator activated receptor gamma
                     2"
                     /protein_id="AAK17925.1"
                     /db_xref="GI:13274398"
                     /translation="M"
ORIGIN

  Query Match           90.5%;  Score 455.2;  DB 9;  Length 984;
  Best Local Similarity  96.8%;  Pred. No. 7.4e-97;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 421

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||| |||||||          ||||||||||||
Db        422 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 474

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        475 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 534

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||||| |||||||||| |||||||||||||||||
Db        535 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 594

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        595 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 654

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| |||||||||||||| ||||||||||||||||||||||||||||| |||||||||
Db        655 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 714

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        715 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 774

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 834

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              ||||||||||||||||||||||||
Db        835 CCAAATATTTGGAAACTGATGTCT 858


RESULT 3
AR121474
LOCUS       AR121474                1100 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 10 from patent US 6159734.
ACCESSION   AR121474
VERSION     AR121474.1  GI:14105050
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   McKay,R., Borchers,A.H. and Baker,B.F.
  TITLE     Antisense modulation of peroxisome proliferator-activated receptor
            gamma expression
  JOURNAL   Patent: US 6159734-A 10 12-DEC-2000;
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match           90.5%;  Score 455.2;  DB 6;  Length 1100;
  Best Local Similarity  96.8%;  Pred. No. 7.3e-97;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;
```

```
Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 298

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||| |||||||           ||||||||||
Db        299 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 351

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 411

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db        412 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 471

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        472 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 531

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| |||||||||||||| ||||||||||||||||||||||||||||| |||||||||
Db        532 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 591

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        592 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 651

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 711

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db        712 CCAAATATTTGGAAACTGATGTCT 735
```

```
RESULT 4
AB005520
LOCUS       AB005520               1100 bp    DNA     linear   PRI 10-NOV-1997
DEFINITION  Homo sapiens ppar gamma2 gene for peroxisome proliferator
            activated-receptor gamma, partial cds and 5' flanking.
ACCESSION   AB005520
VERSION     AB005520.1  GI:2605488
KEYWORDS    ppar gamma2; peroxisome proliferator activated-receptor gamma.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Okazawa,H., Mori,H., Tamori,Y., Araki,S., Niki,T., Masugi,J.,
            Kawanishi,M., Kubota,T., Sinoda,H. and Kasuga,M.
  TITLE     No coding mutations are detected in the peroxisome
            proliferator-activated receptor- gene in Japanese patients with
            lipoatrophic diabetes
  JOURNAL   Diabetes (1997) In press
REFERENCE   2  (bases 1 to 1100)
  AUTHORS   Okazawa,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1997) Hideki Okazawa, Kobe University School of
            Medicine, 2nd Department of Internal Medicine; 7-5-1 Kusunoki-cho
            chuo-ku, Kobe 650, Japan (E-mail:okazawa@med.kobe-u.ac.jp,
            Tel:81-78-341-7451, Fax:81-78-382-2080)
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /tissue_type="placenta"
     gene            859. .940
                     /gene="ppar gamma2"
     CDS             859. .>940
                     /gene="ppar gamma2"
                     /note="ppar gamma2 specific coding region"
                     /codon_start=1
                     /product="peroxisome proliferator activated-receptor
                     gamma"
                     /protein_id="BAA23353.1"
                     /db_xref="GI:2605489"
                     /translation="MGETLGDSPIDPESDSFTDTLSANISQ"
     exon            <859. .940
                     /gene="ppar gamma2"
ORIGIN
```

```
  Query Match           90.5%;  Score 455.2;  DB 9;  Length 1100;
  Best Local Similarity   96.8%;  Pred. No. 7.3e-97;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 298

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||| |||||||           ||||||||||
Db        299 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 351

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 411

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db        412 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 471

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        472 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 531

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| |||||||||||||| ||||||||||||||||||||||||||||| |||||||||
Db        532 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 591

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        592 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 651

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 711

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db        712 CCAAATATTTGGAAACTGATGTCT 735
```

```
RESULT 5
AX951686/c
LOCUS       AX951686               85873 bp    DNA     linear   PAT 09-JAN-2004
DEFINITION  Sequence 3 from Patent WO03093310.
ACCESSION   AX951686
VERSION     AX951686.1  GI:40782040
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Bullerdiek,J.
  TITLE     Novel nucleic acid sequences and proteins of tumors and neoplasias
            of the thyroid gland
  JOURNAL   Patent: WO 03093310-A 3 13-NOV-2003;
            Universitaet Bremen (DE)
FEATURES             Location/Qualifiers
     source          1. .85873
                     /organism="Homo sapiens"
```

```
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             90.5%;  Score 455.2;  DB 6;  Length 85873;
  Best Local Similarity   96.8%;  Pred. No. 4.7e-97;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54472 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 54413

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              ||||||||||||||||||||||||||||||||| |||||||          ||||||||||
Db      54412 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 54360

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54359 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 54300

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db      54299 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 54240

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      54239 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 54180

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| |||||||||||||| |||||||||||||||||||||||||||| |||||||||
Db      54179 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 54120

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      54119 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 54060

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54059 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 54000

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db      53999 CCAAATATTTGGAAACTGATGTCT 53976


RESULT 6
AC093174
LOCUS       AC093174               135675 bp    DNA     linear   PRI 08-NOV-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-167M22 map 3p, complete
            sequence.
ACCESSION   AC093174
VERSION     AC093174.2  GI:24796717
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 135675)
  AUTHORS   Wu,Q., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J.,
            Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D.,
            Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C.,
            Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B.,
            Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Chromosome 3p genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 135675)
  AUTHORS   Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2001) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
REFERENCE   3  (bases 1 to 135675)
  AUTHORS   Wu,Q., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J.,
            Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D.,
            Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C.,
            Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B.,
            Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y.,
            Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X.,
            Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and
            Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
COMMENT     On Nov 8, 2002 this sequence version replaced gi:15148930.
            --------------Genome Center
            Center:Beijing Center
            Center code:Beijing
            Website:http://hgc.igtp.ac.cn
            http://www.genomics.org.cn
            Contact:hgc@igtp.ac.cn
            -------------- Project Information
            Center project name:1% project
            Center clone name: RP11-167M22
            -------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator: ET 55% of reads
            Chemistry: Dye-terminator Big Dye; 45% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 0 bases at least Q40
            Consensus quality: 0 bases at least Q30
            Consensus quality: 6 bases at least Q20
            Insert size: 3392; sum-of-contigs
            Quality coverage:   0.00x in Q20 bases;sum-of-contigs
            --------------.
FEATURES             Location/Qualifiers
     source          1. .135675
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3p"
                     /clone="RP11-167M22"
ORIGIN

  Query Match             90.5%;  Score 455.2;  DB 9;  Length 135675;
  Best Local Similarity   96.8%;  Pred. No. 4.5e-97;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81198 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 81257

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              ||||||||||||||||||||||||||||||||| |||||||          ||||||||||
Db      81258 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 81310
```

```
Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81311 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 81370

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              ||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db      81371 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 81430

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      81431 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 81490

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| |||||||||||||| |||||||||||||||||||||||||||| |||||||||
Db      81491 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 81550

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      81551 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 81610

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81611 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 81670

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db      81671 CCAAATATTTGGAAACTGATGTCT 81694
```

```
RESULT 7
AY157024
LOCUS       AY157024              148828 bp    DNA     linear   PRI 16-JUN-2004
DEFINITION  Homo sapiens peroxisome proliferative activated receptor gamma
            (PPARG) gene, complete cds.
ACCESSION   AY157024
VERSION     AY157024.2  GI:48762804
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148828)
  AUTHORS   Rieder,M.J., da Ponte,S.H., Kuldanek,S.A., Rajkumar,N., Smith,J.D.,
            Toth,E.J. and Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2002) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
REFERENCE   2  (bases 1 to 148828)
  AUTHORS   Rieder,M.J., Daniels,R.L., da Ponte,S.H., Hastings,N.C.,
            Ahearn,M.O., Rajkumar,N., Yi,Q. and Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-2004) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
  REMARK    Sequence update by submitter
COMMENT     On Jun 16, 2004 this sequence version replaced gi:23953882.
            To cite this work please use:  SeattleSNPs. NHLBI Program for
            Genomic Applications, UW-FHCRC, Seattle, WA
            (URL:http://pga.mbt.washington.edu).
FEATURES             Location/Qualifiers
     source          1. .148828
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     repeat_region   9. .59
                     /rpt_family="L1"
                     /rpt_type=dispersed
     variation       54
                     /frequency="0.12"
                     /replace="g"
     repeat_region   60. .351
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       147
                     /frequency="0.03"
                     /replace="t"
     variation       252
                     /frequency="0.03"
                     /replace="g"
     repeat_region   352. .431
                     /rpt_family="L1"
                     /rpt_type=dispersed
     variation       374
                     /frequency="0.05"
                     /replace="c"
     repeat_region   435. .759
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       1555
                     /frequency="0.02"
                     /replace="t"
     misc_feature    1754. .1925
                     /note="Region not scanned for variation"
     repeat_region   1784. .2127
                     /rpt_family="L1"
                     /rpt_type=dispersed
     repeat_region   2032. .2632
                     /rpt_family="L1"
                     /rpt_type=dispersed
     gene            2048. .148443
                     /gene="PPARG"
     mRNA            join(2048. .2143,26500. .26573,93809. .94036,95427. .95596,
                     106715. .106853,119980. .120179,130803. .131253,
                     147997. .148443)
                     /gene="PPARG"
                     /product="peroxisome proliferative activated receptor
                     gamma"
     variation       2076
                     /gene="PPARG"
                     /frequency="0.15"
                     /replace="t"
     variation       2107
                     /gene="PPARG"
                     /frequency="0.01"
                     /replace="a"
     variation       2186
                     /gene="PPARG"
                     /frequency="0.22"
                     /replace="c"
     variation       2382
                     /gene="PPARG"
                     /frequency="0.01"
                     /replace="t"
     variation       2407
                     /gene="PPARG"
                     /frequency="0.05"
                     /replace="t"
     variation       2499
                     /gene="PPARG"
                     /frequency="0.04"
                     /replace="t"
     variation       2767
                     /gene="PPARG"
                     /frequency="0.07"
                     /replace="g"
     variation       3035
                     /gene="PPARG"
                     /frequency="0.01"
                     /replace="a"
     variation       3198
                     /gene="PPARG"
                     /frequency="0.04"
                     /replace="t"
     variation       3354
                     /gene="PPARG"
                     /frequency="0.01"
                     /replace=""
```

```
variation       4160
                /gene="PPARG"
                /frequency="0.01"
                /replace="c"
misc_feature    4220. .25559
                /gene="PPARG"
                /note="Region not scanned for variation"
repeat_region   5952. .6004
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   8735. .9019
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   9637. .9935
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   10092. .10404
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   10423. .10537
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   11193. .11356
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   12140. .12449
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   12949. .13255
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   13277. .13383
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   16397. .16653
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   17063. .17196
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   17325. .17622
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   17653. .17733
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   17766. .17984
                /rpt_family="CR1"
                /rpt_type=dispersed
repeat_region   18787. .18845
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   19727. .19804
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   19757. .19813
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   20083. .20181
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   20182. .20524
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   20525. .20660
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   20937. .21239
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   21961. .22204
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   22205. .22492
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   22493. .22581
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   22582. .22916
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   22917. .23104
                /rpt_family="L2"
                /rpt_type=dispersed
```

Query Match 90.5%; Score 455.2; DB 9; Length 148828;
Best Local Similarity 96.8%; Pred. No. 4.4e-97;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;

```
Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65093 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 65152

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||||| |||||||       ||||||||||
Db      65153 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 65205

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65206 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 65265

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db      65266 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 65325

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65326 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 65385

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| |||||||||||||| |||||||||||||||||||||||||||| |||||||||
Db      65386 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 65445

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      65446 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 65505

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65506 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 65565

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||
Db      65566 CCAAATATTTGGAAACTGATGTCT 65589
```

```
RESULT 8
AC090947/c
LOCUS       AC090947              166049 bp    DNA     linear   PRI 08-NOV-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-30G23 map 3p, complete
            sequence.
ACCESSION   AC090947 AC016333
VERSION     AC090947.2  GI:24796728
KEYWORDS    HTG; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166049)
  AUTHORS   Wu,Q., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J.,
            Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D.,
            Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C.,
            Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B.,
            Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
```

```
            Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
            Wang,Y., Wu,D., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Chromosome 3p genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166049)
  AUTHORS   Wu,Q., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J.,
            Ding,H., Dong,W., Pan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D.,
            Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C.,
            Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B.,
            Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
            Wang,Y., Wu,D., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2001) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
REFERENCE   3  (bases 1 to 166049)
  AUTHORS   Wu,Q., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J.,
            Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D.,
            Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C.,
            Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B.,
            Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y.,
            Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X.,
            Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and
            Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
COMMENT     On Nov 8, 2002 this sequence version replaced gi:13384351.
            --------------Genome Center
            Center:Beijing Center
            Center code:Beijing
            Website:http://hgc.igtp.ac.cn
            http://www.genomics.org.cn
            Contact:hgc@igtp.ac.cn
            -------------- Project Information
            Center project name:1% project
            Center clone name: RP11-30G23
            -------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator: ET 55% of reads
            Chemistry: Dye-terminator Big Dye; 45% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 499 bases at least Q40
            Consensus quality: 605 bases at least Q30
            Consensus quality: 674 bases at least Q20
            Insert size: 692; sum-of-contigs
            Quality coverage:   2.80x in Q20 bases;sum-of-contigs
            --------------.
FEATURES             Location/Qualifiers
     source          1. .166049
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3p"
                     /clone="RP11-30G23"
ORIGIN

  Query Match           90.5%;  Score 455.2;  DB 9;  Length 166049;
  Best Local Similarity  96.8%;  Pred. No. 4.4e-97;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145789 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 145730

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||||||| ||||||| ||||||||||
Db     145729 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 145677

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145676 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 145617

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db     145616 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 145557

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     145556 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 145497

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| ||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db     145496 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 145437

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     145436 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 145377

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145376 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 145317

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||
Db     145316 CCAAATATTTGGAAACTGATGTCT 145293


RESULT 9
AC151849/c
LOCUS       AC151849              186028 bp    DNA     linear   HTG 08-OCT-2004
DEFINITION  Callithrix jacchus clone CH259-42F14, WORKING DRAFT SEQUENCE, 2
            ordered pieces.
ACCESSION   AC151849
VERSION     AC151849.1  GI:53983836
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Callithrix jacchus (white-tufted-ear marmoset)
  ORGANISM  Callithrix jacchus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
            Callithrix.
REFERENCE   1  (bases 1 to 186028)
  AUTHORS   Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
            Peng,Z., Malinov,I. and Rubin,E.M.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186028)
  AUTHORS   Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
            Peng,Z., Malinov,I. and Rubin,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2004) Genome Sciences, Lawrence Berkeley National
            Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
COMMENT
            Sequence Produced by Berkeley PGA
            Web site: http://pga.lbl.gov
            Center Code: PGABERK
            Center Project Name: J108-42F14
            Bac Clone Name: CH259-42F14

            This sequence has been compared to sequences of other species
            using Vista (http://www-gsd.lbl.gov/VISTA). The results can be
            viewed at:
            http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=PPARG
```

```
          The order-orientation of the draft sequence was accomplished by
          using:
          Avid (http://baboon.math.berkeley.edu/mavid) and paired end
          information.

          Funding agent: Programs for Genomic Applications (NHLBI)

          Summary Statistics:
          Sequencing vector: Plasmid; pUC18
          Chemistry: Dye-terminator Big Dye
          Assembly program: Phrap version 0.990329.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1    26679: contig of 26679 bp in length
          *   26680    26779: gap of unknown length
          *   26780   186028: contig of 159249 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186028
                     /organism="Callithrix jacchus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9483"
                     /clone="CH259-42F14"
ORIGIN

  Query Match             72.1%;  Score 362.8;  DB 2;  Length 186028;
  Best Local Similarity   91.1%;  Pred. No. 2.5e-75;
  Matches  468;  Conservative    0;  Mismatches   27;  Indels   19;  Gaps    7;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              |||| ||| |||||||||||||||||||||||||||||||||||||||||| |||||||||
Db      84782 TAGAATAAATACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAATGTTTAAACT 84723

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||| |||||||||||||||||       ||||||||||||
Db      84722 ATGGATACATATTTGAATTCCAAAGTTTTCTTCAGATAATGT-------GATTAGAGATT 84670

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      84669 CAACCAGGAATAGACACCAAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 84610

Qy        181 GCAAGAGATTT-AAGTTTTCCATTTAAGAAGCCAT---TGTGAATTATACAACAATAAAA 236
              || |||||||| |||||||||||||||| |||| ||    |||||||| ||  ||||| |||
Db      84609 GCGAGAGATTTAAAGTTTTCCATTTAATAAGCAATTGATGTGAATTTTATGACAAT-AAA 84551

Qy        237 AATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACG 296
              |||||||  |||||||||||||||  |||| ||||||||||||||| ||||||||||||
Db      84550 AATGCAAGGCGATATTGAACAGTCTCCGCTCTTATAATTCTAAATACTGTACAGTTCACG 84491

Qy        297 CCCCTCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTA---TC 352
              |||||||| ||||||| | ||||| |||||| ||||||||||||||||||||||   ||
Db      84490 CCCCTCACAAGACACTCAGCATGTGGGTCACTGGCGAGACAGTGTGGCAATATTATTTTC 84431

Qy        353 CCTGTAATGTACCAAGTCTTGCCAGAGCAGTG---AACATTATGACACAACTTTTTGTCA 409
              |||||||||||||||||||||| |||||||   ||||||||||||||| ||||||||||
Db      84430 CCTGTAATGTACCAAGTCTTGCCAAAGCAGTGAACAACATTATGACACAACATTTTGTCA 84371

Qy        410 CAGCTGGCTCCTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATT 469
              |||||||||||||| |||||||||||||||||||| ||||||||||||||||||||||||
Db      84370 GAGCTGGCTCCTAACAGGACAGTGCCAGCCAATTCAGGCCCAGTCCTTTCTGTGTTTATT 84311

Qy        470 CCCATCTCTCCCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||||||||||
Db      84310 CCCATCTCTCCCAAATATTTGGAAACTGATGTCT 84277
```

```
RESULT 10
AC136055/c
LOCUS       AC136055              191540 bp    DNA     linear   HTG 29-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-123L10, *** SEQUENCING IN PROGRESS
            ***, 62 unordered pieces.
ACCESSION   AC136055
VERSION     AC136055.1  GI:24417926
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 191540)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191540)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
```

```
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: KDDG
    Center clone name: CH230-123L10
-------------- Summary Statistics
    Sequencing vector: Plasmid;
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 138097 bases at least Q40
    Consensus quality: 143025 bases at least Q30
    Consensus quality: 147154 bases at least Q20
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1135: contig of 1135 bp in length
*     1136     1235: gap of unknown length
*     1236     2684: contig of 1449 bp in length
*     2685     2784: gap of unknown length
*     2785     4087: contig of 1303 bp in length
*     4088     4187: gap of unknown length
*     4188     5601: contig of 1414 bp in length
*     5602     5701: gap of unknown length
*     5702     7184: contig of 1483 bp in length
*     7185     7284: gap of unknown length
*     7285     8589: contig of 1305 bp in length
*     8590     8689: gap of unknown length
*     8690    10203: contig of 1514 bp in length
*    10204    10303: gap of unknown length
*    10304    11907: contig of 1604 bp in length
*    11908    12007: gap of unknown length
*    12008    13032: contig of 1025 bp in length
*    13033    13132: gap of unknown length
*    13133    14692: contig of 1560 bp in length
*    14693    14792: gap of unknown length
*    14793    15997: contig of 1205 bp in length
*    15998    16097: gap of unknown length
*    16098    17167: contig of 1070 bp in length
*    17168    17267: gap of unknown length
*    17268    18690: contig of 1423 bp in length
*    18691    18790: gap of unknown length
*    18791    20652: contig of 1862 bp in length
*    20653    20752: gap of unknown length
*    20753    22248: contig of 1496 bp in length
*    22249    22348: gap of unknown length
*    22349    24000: contig of 1652 bp in length
*    24001    24100: gap of unknown length
*    24101    25384: contig of 1284 bp in length
*    25385    25484: gap of unknown length
*    25485    26797: contig of 1313 bp in length
*    26798    26897: gap of unknown length
*    26898    28389: contig of 1492 bp in length
*    28390    28489: gap of unknown length
*    28490    29812: contig of 1323 bp in length
*    29813    29912: gap of unknown length
*    29913    31236: contig of 1324 bp in length
*    31237    31336: gap of unknown length
*    31337    33116: contig of 1780 bp in length
*    33117    33216: gap of unknown length
*    33217    34328: contig of 1112 bp in length
*    34329    34428: gap of unknown length
*    34429    37011: contig of 2583 bp in length
*    37012    37111: gap of unknown length
*    37112    39760: contig of 2649 bp in length
*    39761    39860: gap of unknown length
*    39861    41317: contig of 1457 bp in length
*    41318    41417: gap of unknown length
*    41418    44017: contig of 2600 bp in length
*    44018    44117: gap of unknown length
*    44118    45721: contig of 1604 bp in length
*    45722    45821: gap of unknown length
*    45822    47208: contig of 1387 bp in length
*    47209    47308: gap of unknown length
*    47309    49161: contig of 1853 bp in length
*    49162    49261: gap of unknown length
*    49262    52550: contig of 3289 bp in length
*    52551    52650: gap of unknown length
*    52651    54753: contig of 2103 bp in length
*    54754    54853: gap of unknown length
*    54854    56484: contig of 1631 bp in length
*    56485    56584: gap of unknown length
*    56585    58264: contig of 1680 bp in length
*    58265    58364: gap of unknown length
*    58365    61440: contig of 3076 bp in length
*    61441    61540: gap of unknown length
*    61541    63616: contig of 2076 bp in length
*    63617    63716: gap of unknown length
*    63717    65642: contig of 1926 bp in length
*    65643    65742: gap of unknown length
*    65743    68360: contig of 2618 bp in length
*    68361    68460: gap of unknown length
*    68461    70926: contig of 2466 bp in length
*    70927    71026: gap of unknown length
*    71027    74092: contig of 3066 bp in length
*    74093    74192: gap of unknown length
*    74193    76455: contig of 2263 bp in length
*    76456    76555: gap of unknown length
*    76556    79002: contig of 2447 bp in length
*    79003    79102: gap of unknown length
*    79103    82482: contig of 3380 bp in length
*    82483    82582: gap of unknown length
*    82583    85503: contig of 2921 bp in length
*    85504    85603: gap of unknown length
*    85604    87769: contig of 2166 bp in length
*    87770    87869: gap of unknown length
*    87870    90671: contig of 2802 bp in length
*    90672    90771: gap of unknown length
*    90772    94059: contig of 3288 bp in length
*    94060    94159: gap of unknown length
*    94160    98678: contig of 4519 bp in length
*    98679    98778: gap of unknown length
*    98779   104079: contig of 5301 bp in length
*   104080   104179: gap of unknown length
*   104180   108047: contig of 3868 bp in length
*   108048   108147: gap of unknown length
*   108148   112386: contig of 4239 bp in length
*   112387   112486: gap of unknown length
*   112487   116184: contig of 3698 bp in length
*   116185   116284: gap of unknown length
*   116286   122755: contig of 6471 bp in length
*   122756   122855: gap of unknown length

  Query Match             43.7%;  Score 219.6;  DB 2;  Length 191540;
  Best Local Similarity   75.3%;  Pred. No. 1.3e-41;
  Matches  342;  Conservative    0;  Mismatches   99;  Indels   13;  Gaps    5;

Qy         57 AACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGA 116
              ||  ||   || ||||||||||| | | | | ||||||||||||||||||||      |||
Db      69364 AAGAATATGTATATATTTGAATACAAGAGTATTCTTCAGATAATGTGATTA------AGA 69311

Qy        117 GATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTT-TCTGGTATTT 175
              |||||||| |||| || ||  ||  |||| ||| |||||| || || ||| || |||
Db      69310 ATTTCAACCAAGGATGGATACTTAAGAAAAATGTTGGCCAAATGAGTTTATCTAGTGTTT 69251

Qy        176 CATAAGCAAGAGATTTAAGTTTTCCATTTAAGAA--GCCATTGTGAATTATACAACAATA 233
              |||||   | ||||| |||||||| |||||| ||   |  | |||  |  |   |    |
Db      69250 CATAACTTAAAGATTAAAGTTTTCTATTTAAAAAAGCCACTGGTGTGTATTTTACTGCAA 69191
```

```
Qy        234 AAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTC 293
              |||     | |  ||||| |||||||||| ||||| |||||| |||||| |||||||||
Db      69190 TTAAAAAGCAATCAATATTAAACAGTCTCTGCTCTGGTAATTCCAAATACTGTACAGTTC 69131

Qy        294 ACGCCCCTCACGAGACACTGAACATGTGGTCACCGGCGAGACAGTGTGGCAATATTATCC 353
              || |||||||| |       |       |||||| |||||||||||| ||||  || |||
Db      69130 ACACCCCTCACAAACACTGAATGTGTGGGTCACTGGCGAGACAGTGTAGCAACGTTTTCC 69071

Qy        354 CTGTAATGTACCAAGTCTTGCCA---GAGCAGTGAACATTATGACACAACTTTTTGTCAC 410
              ||||||||||||||||||||    |||||  | |||||||||||| | |||||||||
Db      69070 TTGTAATGTACCAAGTCTTGCCAAAGAAGCAGACAGCATTATGACACACCATTTTGTCAC 69011

Qy        411 AGCTGGCTCCTAAT-AGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATT 469
              |||||||||  ||| |||||||||||||||||||| |||   |||||||||||||||||
Db      69010 AGCTGGCTCTCAATCAGGACAGTGCCAGCCAATTCAGGCCTGATCCTTTCTGTGTTTATT 68951

Qy        470 CCCATCTCTCCCAAATATTTGGAAACTGATGTCT 503
              |||| |||||||||||||||| |||||| |||||
Db      68950 CCCACCTCTCCCAAATATTTGAAAACTGGTGTCT 68917
```

```
RESULT 11
AC091418/c
LOCUS       AC091418               245724 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-2G6, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC091418
VERSION     AC091418.5  GI:24941354
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 245724)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245724)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 245724)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:22855703.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: TUQY
                Center clone name: CH230-2G6
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 220652 bases at least Q40
                Consensus quality: 223895 bases at least Q30
                Consensus quality: 225929 bases at least Q20
                Estimated insert size: 229375; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   245724: contig of 245724 bp in length.
FEATURES              Location/Qualifiers
     source           1. .245724
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10116"
                      /clone="CH230-2G6"
     misc_feature     1. .1001
                      /note="wgs_contig"
     misc_feature     243338. .245724
                      /note="wgs_contig"
ORIGIN

  Query Match             43.7%;  Score 219.6;  DB 2;  Length 245724;
  Best Local Similarity   75.3%;  Pred. No. 1.3e-41;
  Matches  342;  Conservative    0;  Mismatches   99;  Indels   13;  Gaps    5;

Qy         57 AACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGA 116
```

```
Db      15463 AAGAATATGTATATATTTGAATACAAGAGTATTCTTCAGATAATGTGATTA------AGA 15410

Qy        117 GATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTT-TCTGGTATTT 175

Db      15409 ATTTCAACCAAGGATGGATACTTAAGAAAAATGTTGGCCAAATGAGTTTATCTAGTGTTT 15350

Qy        176 CATAAGCAAGAGATTTAAGTTTTCCATTTAAGAA--GCCATTGTGAATTATACAACAATA 233

Db      15349 CATAACTTAAAGATTAAAGTTTTCTATTTAAAAAAGCCACTGGTGTGTATTTTACTGCAA 15290

Qy        234 AAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTC 293

Db      15289 TTAAAAAGCAATCAATATTAAACAGTCTCTGCTCTGGTAATTCCAAATACTGTACAGTTC 15230

Qy        294 ACGCCCCTCACGAGACACTGAACATGTGGTCACCGGCGAGACAGTGTGGCAATATTATCC 353

Db      15229 ACACCCCTCACAAACACTGAATGTGTGGGTCACTGGCGAGACAGTGTAGCAACGTTTTCC 15170

Qy        354 CTGTAATGTACCAAGTCTTGCCA---GAGCAGTGAACATTATGACACAACTTTTTGTCAC 410

Db      15169 TTGTAATGTACCAAGTCTTGCCAAAGAAGCAGACAGCATTATGACACACCATTTTGTCAC 15110

Qy        411 AGCTGGCTCCTAAT-AGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATT 469

Db      15109 AGCTGGCTCTCAATCAGGACAGTGCCAGCCAATTCAGGCCTGATCCTTTCTGTGTTTATT 15050

Qy        470 CCCATCTCTCCCAAATATTTGGAAACTGATGTCT 503

Db      15049 CCCACCTCTCCCAAATATTTGAAAACTGGTGTCT 15016
```

RESULT 12
AC120668/c
LOCUS AC120668 250169 bp DNA linear HTG 23-NOV-2002
DEFINITION Rattus norvegicus clone CH230-24K10, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.
ACCESSION AC120668
VERSION AC120668.4 GI:25188255
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 250169)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 250169)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 250169)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 23, 2002 this sequence version replaced gi:23322270.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GXLO
Center clone name: CH230-24K10
-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 225414 bases at least Q40
Consensus quality: 228258 bases at least Q30
Consensus quality: 229994 bases at least Q20
Estimated insert size: 226717; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     9799: contig of 9799 bp in length
            *      9800     9899: gap of unknown length
            *      9900   248957: contig of 239058 bp in length
            *    248958   249057: gap of unknown length
            *    249058   250169: contig of 1112 bp in length.
FEATURES             Location/Qualifiers
     source          1. .250169
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-24K10"
     misc_feature    4754. .5526
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH280541"
     misc_feature    9900. .10968
                     /note="wgs_contig"
     misc_feature    11406. .14106
                     /note="wgs_contig"
     misc_feature    complement(234439. .235077)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH280540"
     misc_feature    245037. .246234
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    247156. .248957
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN

  Query Match             43.7%;  Score 219.6;  DB 2;  Length 250169;
  Best Local Similarity   75.3%;  Pred. No. 1.3e-41;
  Matches  342;  Conservative    0;  Mismatches   99;  Indels   13;  Gaps    5;

Qy         57 AACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGA 116
              ||  ||   || ||||||||| | | | | |||||||||||||||||||      |||
Db     154629 AAGAATATGTATATATTTGAATACAAGAGTATTCTTCAGATAATGTGATTA------AGA 154576

Qy        117 GATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTT-TCTGGTATTT 175
              |||||||| |||| || ||  ||  |||| ||| |||||| || || ||| || |||
Db     154575 ATTTCAACCAAGGATGGATACTTAAGAAAAATGTTGGCCAAATGAGTTTATCTAGTGTTT 154516

Qy        176 CATAAGCAAGAGATTTAAGTTTTCCATTTAAGAA--GCCATTGTGAATTATACAACAATA 233
              |||||   | ||||| |||||||| |||||| ||   |  | |||  |  |  |     |
Db     154515 CATAACTTAAAGATTAAAGTTTTCTATTTAAAAAAGCCACTGGTGTGTATTTTACTGCAA 154456

Qy        234 AAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTC 293
              |||     | |  ||||| |||||||||| ||||| |||||| |||||| |||||||||
Db     154455 TTAAAAAGCAATCAATATTAAACAGTCTCTGCTCTGGTAATTCCAAATACTGTACAGTTC 154396

Qy        294 ACGCCCCTCACGAGACACTGAACATGTGGTCACCGGCGAGACAGTGTGGCAATATTATCC 353
              || |||||||| |       |      |||||| ||||||||||||| ||||  || |||
Db     154395 ACACCCCTCACAAACACTGAATGTGTGGGTCACTGGCGAGACAGTGTAGCAACGTTTTCC 154336

Qy        354 CTGTAATGTACCAAGTCTTGCCA---GAGCAGTGAACATTATGACACAACTTTTTGTCAC 410
              |||||||||||||||||||||    |||||  | ||||||||||| | |||||||||
Db     154335 TTGTAATGTACCAAGTCTTGCCAAAGAAGCAGACAGCATTATGACACACCATTTTGTCAC 154276

Qy        411 AGCTGGCTCCTAAT-AGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATT 469
              |||||||||  ||| ||||||||||||||||||| |||   |||||||||||||||||
Db     154275 AGCTGGCTCTCAATCAGGACAGTGCCAGCCAATTCAGGCCTGATCCTTTCTGTGTTTATT 154216

Qy        470 CCCATCTCTCCCAAATATTTGGAAACTGATGTCT 503
              |||| |||||||||||||||| |||||| |||||
Db     154215 CCCACCTCTCCCAAATATTTGAAAACTGGTGTCT 154182


RESULT 13
S79407
LOCUS       S79407                   682 bp    DNA     linear   ROD 09-DEC-1995
DEFINITION  mPPAR gamma 2=peroxisome proliferator-activated receptor gamma
            isoform 2 (promoter) [mice, Genomic, 682 nt].
ACCESSION   S79407
VERSION     S79407.1  GI:1110564
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   Zhu,Y., Qi,C., Korenberg,J.R., Chen,X.N., Noya,D., Rao,M.S. and
            Reddy,J.K.
  TITLE     Structural organization of mouse peroxisome proliferator-activated
            receptor gamma (mPPAR gamma) gene: alternative promoter use and
            different splicing yield two mPPAR gamma isoforms
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7921-7925 (1995)
  MEDLINE   95372391
   PUBMED   7644514
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 170483] from the original journal article.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Mus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10095"
     mRNA            616. .>682
     gene            680. .682
                     /gene="mPPAR<gamma>2"
                     /note="peroxisome proliferator-activated receptor gamma
                     isoform 2"
ORIGIN

  Query Match             43.0%;  Score 216.4;  DB 10;  Length 682;
  Best Local Similarity   77.6%;  Pred. No. 1.3e-40;
  Matches  354;  Conservative    0;  Mismatches   86;  Indels   16;  Gaps    7;

Qy         57 AACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGA 116
              ||  |||  || || ||||| | | | | | ||||||||||  |||||||||      ||
Db        108 AAGAATGTGTATATGTTTGAGTACAAGAATATTCTTCAGAT-GTGTGATTAG------GA 160

Qy        117 GATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTT-TCTGGTATTT 175
              | ||||||||  |||| | ||  ||  |||||||||| ||||||| | || |||||| |||
Db        161 GTTTCAACCAAAGATAAATACTTAAGAAAAACTTTGGCCAAATACGTTTATCTGGTGTTT 220

Qy        176 CATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCA---TTGTGAATTATACAACAAT 232
              |||||   ||||||| | |||||| |||| | ||||||     |||| ||| |||  ||||
Db        221 CATAACTTAGAGATTAAGGTTTTCTATTTTAAAAGCCACTGGTGTGTATTTTACTGCAAT 280

Qy        233 AAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTT 292
              |||    | |  |||||||||| ||||| ||||| |||||| || ||| ||||||||
Db        281 TTTAAAAAGCAATCAATATTGAACAATCTCTGCTCTGGTAATTCCAACTACTGTACAGTT 340

Qy        293 CACGCCCCTCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTAT 351
              ||||||||||||   ||| ||||  ||| |||||| ||||||||| ||| ||||   || |
Db        341 CACGCCCCTCACAGAACAGTGAATGTGTGGGTCACTGGCGAGACAATGTAGCAACGTTTT 400

Qy        352 CCCTGTAATGTACCAAGTCTTGCCA---GAGCAGTGAACATTATGACACAACTTTTTGTC 408
              || |||||||||||||||||||||    |||||  | ||||||||||| | |||||||
Db        401 CCTTGTAATGTACCAAGTCTTGCCAAAGCAGCAGACAGCATTATGACACACCATTTTGTC 460

Qy        409 ACAGCTGGCTCCTAAT-AGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTA 467
              ||| |||||||  | | |||||||||||||||||||| |||   | ||||||||||||
Db        461 ACAACTGGCTCTCAGTCAGGACAGTGCCAGCCAATTCAGGCCTGATTCTTTCTGTGTTTA 520

Qy        468 TTCCCATCTCTCCCAAATATTTGGAAACTGATGTCT 503
              |||||| |||||||||||||||| |||||| |||||
Db        521 TTCCCACCTCTCCCAAATATTTGAAAACTGGTGTCT 556


RESULT 14
AY243584
```

```
LOCUS       AY243584               1159 bp    DNA     linear   ROD 01-JUL-2003
DEFINITION  Mus musculus strain 129S1/SvImJ peroxisome proliferator-activated
            receptor gamma transcript 2 (Pparg) gene, promoter region.
ACCESSION   AY243584
VERSION     AY243584.1  GI:32395962
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1159)
  AUTHORS   Lyons,M.A., Wittenburg,H., Li,R., Walsh,K.A., Leonard,M.R.,
            Churchill,G.A., Carey,M.C. and Paigen,B.
  TITLE     Lith13 Encompasses Lrpap1, Which is a Likely Genetic Determinant of
            Murine Cholesterol Gallstone Formation
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1159)
  AUTHORS   Lyons,M.A., Wittenburg,H., Walsh,K.A., Carey,M.C. and Paigen,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2003) The Jackson Laboratory, 600 Main Street,
            Bar Harbor, ME 04609, USA
FEATURES             Location/Qualifiers
     source          1. .1159
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="129S1/SvImJ"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /map="52.7 cM"
     gene            1. .>1159
                     /gene="Pparg"
                     /note="peroxisome proliferator-activated receptor gamma
                     transcript 2"
     promoter        1. .1159
                     /gene="Pparg"
ORIGIN

  Query Match             43.0%;  Score 216.4;  DB 10;  Length 1159;
  Best Local Similarity   77.6%;  Pred. No. 1.2e-40;
  Matches  354;  Conservative    0;  Mismatches   86;  Indels   16;  Gaps    7;

Qy         57 AACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGA 116
              ||  |||  || || ||||| | | | | ||||||||||  |||||||||      ||
Db        633 AAGAATGTGTATATGTTTGAGTACAAGAATATTCTTCAGAT-GTGTGATTAG------GA 685

Qy        117 GATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTT-TCTGGTATTT 175
              | ||||||||  |||| | ||  ||  ||||||||| ||||||| | || |||||| |||
Db        686 GTTTCAACCAAAGATAAATACTTAAGAAAAACTTTGGCCAAATACGTTTATCTGGTGTTT 745

Qy        176 CATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCA---TTGTGAATTATACAACAAT 232
              |||||   ||||||| | |||||| |||| | ||||||    |||| ||| |||  ||||
Db        746 CATAACTTAGAGATTAAGGTTTTCTATTTTAAAAGCCACTGGTGTGTATTTTACTGCAAT 805

Qy        233 AAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTT 292
                 |||    | |  |||||||||| ||||| ||||| |||||| || ||| ||||||||
Db        806 TTTAAAAAGCAATCAATATTGAACAATCTCTGCTCTGGTAATTCCAACTACTGTACAGTT 865

Qy        293 CACGCCCCTCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTAT 351
              ||||||||||||   ||| ||||  ||| |||||| ||||||||| ||| ||||  || |
Db        866 CACGCCCCTCACAGAACAGTGAATGTGTGGGTCACTGGCGAGACAATGTAGCAACGTTTT 925

Qy        352 CCCTGTAATGTACCAAGTCTTGCCA---GAGCAGTGAACATTATGACACAACTTTTTGTC 408
              || ||||||||||||||||||||||     |||||  | |||||||||||| | |||||||
Db        926 CCTTGTAATGTACCAAGTCTTGCCAAAGCAGCAGACAGCATTATGACACACCATTTTGTC 985

Qy        409 ACAGCTGGCTCCTAAT-AGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTA 467
              ||| |||||||  | | |||||||||||||||||||||| |||   | ||||||||||||
Db        986 ACAACTGGCTCTCAGTCAGGACAGTGCCAGCCAATTCAGGCCTGATTCTTTCTGTGTTTA 1045

Qy        468 TTCCCATCTCTCCCAAATATTTGGAAACTGATGTCT 503
              |||||| |||||||||||||||| |||||| |||||
Db       1046 TTCCCACCTCTCCCAAATATTTGAAAACTGGTGTCT 1081
```

```
RESULT 15
AY236531
LOCUS       AY236531               1160 bp    DNA     linear   ROD 01-JUL-2003
DEFINITION  Mus musculus strain DBA/2J peroxisome proliferator-activated
            receptor gamma transcript 2 (Pparg) gene, promoter region.
ACCESSION   AY236531
VERSION     AY236531.1  GI:32395932
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1160)
  AUTHORS   Lyons,M.A., Wittenburg,H., Li,R., Walsh,K.A., Leonard,M.R.,
            Korstanje,R., Churchill,G.A., Carey,M.C. and Paigen,B.
  TITLE     Lith6 Encompasses PPARgamma and SLC21A1 Which Are Likely Genetic
            Determinants of Murine Cholesterol Gallstone Formation
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1160)
  AUTHORS   Lyons,M.A., Wittenburg,H., Walsh,K.A., Carey,M.C. and Paigen,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2003) The Jackson Laboratory, 600 Main Street,
            Bar Harbor, ME 04609, USA
FEATURES             Location/Qualifiers
     source          1. .1160
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="DBA/2J"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /map="52.7 cM"
     gene            1. .>1160
                     /gene="Pparg"
                     /note="peroxisome proliferator-activated receptor gamma
                     transcript 2"
     promoter        1. .1160
                     /gene="Pparg"
ORIGIN

  Query Match             43.0%;  Score 216.4;  DB 10;  Length 1160;
  Best Local Similarity   77.6%;  Pred. No. 1.2e-40;
  Matches  354;  Conservative    0;  Mismatches   86;  Indels   16;  Gaps    7;

Qy         57 AACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGA 116
              ||  |||  || || ||||| | | | | ||||||||||  |||||||||      ||
Db        634 AAGAATGTGTATATGTTTGAGTACAAGAATATTCTTCAGAT-GTGTGATTAG------GA 686

Qy        117 GATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTT-TCTGGTATTT 175
              | ||||||||  |||| | ||  ||  ||||||||| ||||||| | || |||||| |||
Db        687 GTTTCAACCAAAGATAAATACTTAAGAAAAACTTTGGCCAAATACGTTTATCTGGTGTTT 746

Qy        176 CATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCA---TTGTGAATTATACAACAAT 232
              |||||   ||||||| | |||||| |||| | ||||||    |||| ||| |||  ||||
Db        747 CATAACTTAGAGATTAAGGTTTTCTATTTTAAAAGCCACTGGTGTGTATTTTACTGCAAT 806

Qy        233 AAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTT 292
                 |||    | |  |||||||||| ||||| ||||| |||||| || ||| ||||||||
Db        807 TTTAAAAAGCAATCAATATTGAACAATCTCTGCTCTGGTAATTCCAACTACTGTACAGTT 866

Qy        293 CACGCCCCTCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTAT 351
              ||||||||||||   ||| ||||  ||| |||||| ||||||||| ||| ||||  || |
Db        867 CACGCCCCTCACAGAACAGTGAATGTGTGGGTCACTGGCGAGACAATGTAGCAACGTTTT 926

Qy        352 CCCTGTAATGTACCAAGTCTTGCCA---GAGCAGTGAACATTATGACACAACTTTTTGTC 408
              || ||||||||||||||||||||||     |||||  | |||||||||||| | |||||||
Db        927 CCTTGTAATGTACCAAGTCTTGCCAAAGCAGCAGACAGCATTATGACACACCATTTTGTC 986

Qy        409 ACAGCTGGCTCCTAAT-AGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTA 467
              ||| |||||||  | | |||||||||||||||||||||| |||   | ||||||||||||
Db        987 ACAACTGGCTCTCAGTCAGGACAGTGCCAGCCAATTCAGGCCTGATTCTTTCTGTGTTTA 1046

Qy        468 TTCCCATCTCTCCCAAATATTTGGAAACTGATGTCT 503
```

```
                ||||||| |||||||||||||||| ||||||| |||||
Db       1047 TTCCCACCTCTCCCAAATATTTGAAAACTGGTGTCT 1082
```

Search completed: November 9, 2005, 02:15:43
Job time : 2644.94 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:46:48 ; Search time 407.412 Seconds
(without alignments)
7308.644 Million cell updates/sec

Title: US-09-463-542-3_COPY_818_1320
Perfect score: 503
Sequence: 1 tagagtaagtaccttaggaa.....atatttggaaactgatgtct 503

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	503	100.0	2045	2	AAX19033	Aax19033 Human PPA
	2	455.2	90.5	1100	5	AAF23651	Aaf23651 Human PPA
c	3	455.2	90.5	85873	10	ADH10008	Adh10008 Human chr
	4	455.2	90.5	158417	13	ADS36461	Ads36461 Human aut
c	5	455.2	90.5	166043	12	ADL08127	Adl08127 Human gen
	6	216.4	43.0	1184	10	ACC79611	Acc79611 Mouse per
	7	52	10.3	177	2	AAX19064	Aax19064 Human PPA
	8	45.2	9.0	200000	12	ADO47191	Ado47191 DNA seque
c	9	44.2	8.8	3162	6	ABQ72637	Abq72637 Human MDD
c	10	43.6	8.7	6215	6	ABL33191	Abl33191 Human imm
c	11	41.8	8.3	99916	6	ADI03931	Adi03931 Human enz
c	12	41.8	8.3	116297	12	ADQ97587	Adq97587 Human can
	13	41.2	8.2	32392	6	ABL56203	Abl56203 AmEPV gen
c	14	40.8	8.1	2000	12	ADJ40655	Adj40655 Plant cDN
c	15	40.8	8.1	109661	12	ADQ97818	Adq97818 Human can
c	16	40.2	8.0	376	4	AAH93354	Aah93354 Human chr
c	17	40.2	8.0	2000	8	ADA71938	Ada71938 Rice gene
c	18	40	8.0	8056	8	ABZ10100	Abz10100 Haematopo
	19	39.4	7.8	6191	6	ABL33216	Abl33216 Human imm
	20	39.4	7.8	6191	6	ABK31306	Abk31306 Signal tr
	21	39.4	7.8	6191	6	ABL70281	Abl70281 Chemicall
	22	39.4	7.8	6191	6	ABN80160	Abn80160 Human che
c	23	39.2	7.8	6308	6	ABL33470	Abl33470 Human imm
c	24	39	7.8	6565	4	AAS46466	Aas46466 Tumour su
c	25	39	7.8	6565	6	ABK31327	Abk31327 Signal tr
c	26	39	7.8	37515	6	ABQ66998	Abq66998 Human ang
	27	38.8	7.7	1696	12	ADI43067	Adi43067 Plant tra
	28	38.8	7.7	1696	12	ADO03207	Ado03207 Corn orth
c	29	38.8	7.7	8056	8	ABZ10246	Abz10246 Haematopo
c	30	38.8	7.7	10945	6	ABK28354	Abk28354 DNA trans
c	31	38.8	7.7	19634	8	ABZ10162	Abz10162 Haematopo
c	32	38.8	7.7	19634	8	ABZ10016	Abz10016 Haematopo
c	33	38.8	7.7	19634	13	ADS89398	Ads89398 Oligonucl
c	34	38.8	7.7	19634	13	ADS89672	Ads89672 Oligonucl
c	35	38.4	7.6	2258	4	ABL21792	Abl21792 Drosophil
c	36	38.2	7.6	4152	6	ABL55640	Abl55640 AmEPV ABC
c	37	38.2	7.6	50000	6	ABL56201	Abl56201 AmEPV gen
c	38	38.2	7.6	96589	9	ADA02708	Ada02708 Human ZFH
c	39	38.2	7.6	96589	10	ADB72446	Adb72446 Human ZFH
c	40	38.2	7.6	96589	10	ADE95956	Ade95956 Human ZFH
	41	38	7.6	2035	13	ADT05376	Adt05376 Haemophil
c	42	38	7.6	349980	13	ADT05649	Adt05649 Haemophil
	43	37.8	7.5	2142	2	AAX84309	Aax84309 Stealth v
	44	37.8	7.5	2144	3	AAZ36928	Aaz36928 Nucleotid
c	45	37.8	7.5	5274	11	ACN92690	Acn92690 Breast ca

ALIGNMENTS

RESULT 1
AAX19033
ID AAX19033 standard; DNA; 2045 BP.
XX
AC AAX19033;
XX
DT 13-MAY-1999 (first entry)
XX
DE Human PPAR-gamma-2 proximal promoter, exon B and intron B.
XX
KW Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;
KW regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;
KW lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;
KW insulin resistance; non-insulin-dependent diabetes mellitus;
KW polycystic ovary syndrome; gastrointestinal tract; Crohn's disease;
KW inflammatory bowel disease; ulcerative colitis; bowel cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9905161-A1.
XX
PD 04-FEB-1999.
XX
PF 24-JUL-1998; 98WO-US015411.
XX
PR 25-JUL-1997; 97US-0053692P.
XX
PA (LIGA-) LIGAND PHARM INC.
PA (INSP) INST PASTEUR.
XX
PI Briggs MR, Saladin RS, Auwerx J, Fajas L;
XX
DR WPI; 1999-142844/12.
XX
PT Newly isolated nucleic acid comprising a control region of a human
PT peroxisome proliferator activated receptor (PPAR) gamma gene - useful for
PT identifying modulators that are useful in treating diseases associated
PT with abnormal levels of human PPAR-gamma gene expression.
XX
PS Claim 10; Page 80; 102pp; English.
XX
CC The present invention describes an isolated, purified or enriched nucleic
CC acid comprising a control region of a human peroxisome proliferator

```
CC   activated receptor gamma (PPAR-gamma) gene. The nucleic acids are useful
CC   for screening for agents capable of modulating the expression of a human
CC   PPAR-gamma gene. These agents (modulators) form pharmaceutical
CC   compositions that are useful for treating diseases associated with
CC   high/low levels of human PPAR-gamma gene expression. The diseases include
CC   obesity, anorexia, cachexia, lipodystrophy, lipomas, liposarcomas,
CC   abnormalities associated with anti-human immunodeficiency virus (HIV)
CC   treatment, insulin resistance, non-insulin-dependent diabetes mellitus
CC   (NIDDM), polycystic ovary syndrome, diseases of the gastrointestinal (GI)
CC   tract, inflammatory bowel disease, Crohn's disease, ulcerative colitis
CC   and bowel cancer. The nucleic acids are useful for studying the role of
CC   the PPAR-gamma gene in various diseases and disorders. The structure of
CC   PPAR-gamma enables genetic studies of PPAR- gamma mutations in humans,
CC   and evaluation of its role in disorders like insulin resistance, NIDDM,
CC   and diseases associated with altered adipose tissue function, like
CC   obesity and lipodystrophic syndromes. The nucleic acids are also useful
CC   for gene therapy and the production of transgenic animals, which are
CC   useful in screening assays. The control regions of the nucleic acids
CC   enable screening for modulators of the human PPAR-gamma gene, which are
CC   useful in designing drugs for treating disorders or diseases associated
CC   with the level of PPAR-gamma gene expression. The present sequence
CC   represents human PPAR-gamma-2 proximal promoter, exon B and intron B
XX
SQ   Sequence 2045 BP; 605 A; 422 C; 351 G; 667 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 503;  DB 2;  Length 2045;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-128;
  Matches  503;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        818 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 877

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 937

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 997

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 1057

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 1117

Qy        301 TCACGAGACACTGAACATGTGGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 TCACGAGACACTGAACATGTGGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAAT 1177

Qy        361 GTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 GTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTCC 1237

Qy        421 TAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1238 TAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTCC 1297

Qy        481 CAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||
Db       1298 CAAATATTTGGAAACTGATGTCT 1320


RESULT 2
AAF23651
ID   AAF23651 standard; cDNA; 1100 BP.
XX
AC   AAF23651;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human PPARgamma N-terminal sequence coding sequence.
XX
KW   Cytostatic; antiinflammatory; antisense oligonucleotide; PPARgamma;
KW   peroxisome proliferator-activated receptor gamma; transcription factor;
KW   nuclear hormone receptor; human; infection; inflammation; tumour; ss.
XX
OS   Homo sapiens.
XX
PN   US6159734-A.
XX
PD   12-DEC-2000.
XX
PF   18-JAN-2000; 2000US-00484345.
XX
PR   18-JAN-2000; 2000US-00484345.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Mckay R,  Baker BF,  Borchers AH;
XX
DR   WPI; 2001-070112/08.
DR   P-PSDB; AAB59840.
XX
PT   Novel antisense compounds capable of modulating expression of peroxisome
PT   proliferator-activated receptor gamma useful for diagnosis, prophylaxis
PT   and treatment of diseases associated with expression of the receptor.
XX
PS   Example 15; Col 49-52; 40pp; English.
XX
CC   Peroxisome proliferator-activated receptors (PPARs) are members of the
CC   nuclear hormone receptor subfamily of transcription factors. The present
CC   invention relates to antisense oligonucleotides (see AAF23652-F23731),
CC   targeted to a nucleic acid molecule encoding human PPARgamma, which
CC   specifically hybridises with and inhibits the expression of human
CC   PPARgamma. The present sequence is the coding sequence for the N-terminal
CC   sequence of human PPARgamma. The PPARgamma antisense oligonucleotides of
CC   the present invention can be used in the diagnosis and treatment of
CC   diseases associated with the expression of PPARgamma, e.g. to prevent or
CC   delay infection, inflammation or tumour formation
XX
SQ   Sequence 1100 BP; 334 A; 215 C; 195 G; 356 T; 0 U; 0 Other;

  Query Match             90.5%;  Score 455.2;  DB 5;  Length 1100;
  Best Local Similarity   96.8%;  Pred. No. 3.1e-115;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 298

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||||||||||| |||||||       ||||||||||
Db        299 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 351

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 411

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||| |||||||||| |||||||||||||||||||
Db        412 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 471

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        472 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 531

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| |||||||||||||| |||||||||||||||||||||||||||| |||||||||
Db        532 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 591

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
```

```
Db        592 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 651

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 711

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db        712 CCAAATATTTGGAAACTGATGTCT 735


RESULT 3
ADH10008/c
ID   ADH10008 standard; DNA; 85873 BP.
XX
AC   ADH10008;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Human chromosome 3p25 DNA fragment.
XX
KW   ds; human; chromosome 2; hyperplasia; tumour; thyroid; cancer;
KW   break point; chromosomal band 2p21-22; DRIP; cytostatic; thyromimetic;
KW   chromosome 7p15.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO2003093310-A1.
XX
PD   13-NOV-2003.
XX
PF   02-MAY-2003; 2003WO-EP004642.
XX
PR   01-MAY-2002; 2002DE-01019413.
PR   14-SEP-2002; 2002DE-01042705.
XX
PA   (UYBR-) UNIV BREMEN.
XX
PI   Bullerdiek J;
XX
DR   WPI; 2003-854480/79.
XX
PT   New nucleic acid from human chromosome 2, useful for treatment and
PT   diagnosis of thyroid disease, especially cancer, also related
PT   polypeptides and modulators.
XX
PS   Claim 1; SEQ ID NO 3; 461pp; German.
XX
CC   This invention describes a novel nucleic acid from human chromosome 2
CC   that shows altered expression in hyperplasia and/or tumours, especially
CC   of the thyroid. The invention describes a method for preparing nucleic
CC   acid that can be detected in thyroid cancers where these contain an
CC   aberration of the break point in chromosomal band 2p21-22. The invention
CC   also describes constructs comprising a fragment of the human DRIP gene
CC   and its splice variants, one of which contains all of exons 1-38 while
CC   the other lacks exons 27 and 28. Some tumours contain a fusion protein of
CC   DRIP, including a small segment from chromosome 3. The products of the
CC   invention have cytostatic and thyromimetic activity and are used for the
CC   inhibition of nucleic acid expression by antisense, ribozyme or RNA
CC   interference (RNAi) methods. This sequence represents a fragment of human
CC   chromosome 3p35 near the region encoding the DRIP protein described in
CC   the disclosure of the invention.
XX
SQ   Sequence 85873 BP; 27329 A; 16422 C; 15988 G; 26134 T; 0 U; 0 Other;

  Query Match             90.5%;  Score 455.2;  DB 10;  Length 85873;
  Best Local Similarity   96.8%;  Pred. No. 1.3e-114;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54472 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 54413

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||||||| ||||||| ||||||||||||
Db      54412 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 54360

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54359 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 54300

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||||||||| |||||||||| ||||||||||||
Db      54299 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 54240

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      54239 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 54180

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| |||||||||||||| |||||||||||||||||||||||||||| |||||||||
Db      54179 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 54120

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      54119 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 54060

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54059 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 54000

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db      53999 CCAAATATTTGGAAACTGATGTCT 53976


RESULT 4
ADS36461
ID   ADS36461 standard; DNA; 158417 BP.
XX
AC   ADS36461;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Human autoimmune disease-related genomic DNA sequence - SEQ ID 1675.
XX
KW   single nucleotide polymorphism detection; SNP detection;
KW   rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW   systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW   thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW   glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW   primary systemic vasculitis; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004083403-A2.
XX
PD   30-SEP-2004.
XX
PF   18-MAR-2004; 2004WO-US008461.
XX
PR   18-MAR-2003; 2003US-0455444P.
PR   25-APR-2003; 2003US-0465241P.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Cargill M,  Begovich AB,  Alexander HC;
XX
DR   WPI; 2004-728480/71.
XX
PT   New isolated nucleic acid molecule comprises at least 8 contiguous
PT   nucleotides where one of the nucleotides is a single nucleotide
PT   polymorphism (SNP), useful for diagnosing or treating autoimmune
PT   diseases, e.g. rheumatoid arthritis.
XX
```

```
PS   Claim 16; SEQ ID NO 1675; 123pp; English.
XX
CC   The invention comprises amino acid and coding sequences containing
CC   genetic polymorphisms associated with an altered risk of developing an
CC   autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC   comprises a method of identifying an individual that has an altered risk
CC   of developing an autoimmune disease, comprising detecting a single
CC   nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC   and protein sequences of the invention are useful for diagnosing and
CC   treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC   diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC   bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC   anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC   myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC   present nucleic acid represents a human autoimmune disease-related
CC   genomic DNA sequence of the invention. NOTE: The present sequence is not
CC   shown in the specification, but has been retrieved from the WIPO website.
XX
SQ   Sequence 158417 BP; 47887 A; 30186 C; 31475 G; 48577 T; 0 U; 292 Other;

  Query Match             90.5%;  Score 455.2;  DB 13;  Length 158417;
  Best Local Similarity   96.8%;  Pred. No. 1.6e-114;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69064 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 69123

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||| |||||||     ||||||||||||
Db      69124 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 69176

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69177 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 69236

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db      69237 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 69296

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      69297 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 69356

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| |||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db      69357 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 69416

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      69417 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 69476

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69477 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 69536

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db      69537 CCAAATATTTGGAAACTGATGTCT 69560


RESULT 5
ADL08127/c
ID   ADL08127 standard; DNA; 166043 BP.
XX
AC   ADL08127;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human gene associated with low HDL-C PPARG.
XX
KW   Human; ds; SNP; single nucleotide polymorphism;
KW   high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;
KW   coronary artery disease; gene.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(145136,C)
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   US2004043389-A1.
XX
PD   04-MAR-2004.
XX
PF   04-SEP-2002; 2002US-00235192.
XX
PR   04-SEP-2002; 2002US-00235192.
XX
PA   (VITI-) VITIVITY INC.
XX
PI   Mccarthy J;
XX
DR   WPI; 2004-214170/20.
XX
PT   Determining whether a subject has, or is at risk of developing, an
PT   abnormally low high density lipoprotein-C (HDL-C) level comprises
PT   detecting an allelic variant of a polymorphic region from any of a set of
PT   27 genes.
XX
PS   Disclosure; SEQ ID NO 46; 37pp; English.
XX
CC   The invention relates to determining whether a subject has, or is at risk
CC   of developing, an abnormally low high density lipoprotein-C (HDL-C) level
CC   comprises determining whether the subject has an allelic variant of a
CC   polymorphic region from any of 27 genes (alleles listed in Table 5 of the
CC   specification). Also included are determining whether a male subject has,
CC   or is at risk of developing, an abnormally low HDL-C level, comprising
CC   determining whether the male subject has an allelic variant of a
CC   polymorphic region listed in Table 5 which is associated with abnormally
CC   low HDL-C levels in males, and determining whether a female subject has,
CC   or is at risk of developing, an abnormally low HDL-C level, comprising
CC   determining whether the female subject has an allelic variant of a
CC   polymorphic region listed in Table 5 which is associated with abnormally
CC   low HDL-C levels in females. The allelic variant in determining whether a
CC   subject has, or is at risk of developing, an abnormally low HDL-C level
CC   is APOA 1 CC, CD14 1 CT, COL5A2 1 GG, EDNRB 1 AG or AA, FABP3 1 CT, GBE1
CC   1 AG or GG, LIPC 5 AA, MTHFR 1 CC, VWF 2 GG, or their complements. The
CC   allelic variant in determining whether a male subject has, or is at risk
CC   of developing, an abnormally low HDL-C level, LRP1 3 CC or CT, PAI2 4 GG,
CC   or PPARG 1 CG, or their complements. The allelic variants are also COL5A2
CC   1 GG, CD14 1 CT or CC, and FABP3 1 CT, in combination, or their
CC   complements. The methods are useful for diagnosing (a predisposition to)
CC   abnormally low levels of low high density lipoprotein-C (HDL-C) in a
CC   subject, The methods are useful in diagnosing (a predisposition to) or
CC   prognosticating diseases and disorders associated with abnormal lipid
CC   levels such as vascular and metabolic diseases, e.g., coronary artery
CC   disease. The present sequence is a human gene containing a SNP (single
CC   nucleotide polymorphism associated with low high density lipoprotein-C
CC   (HDL-C) levels.
XX
SQ   Sequence 166043 BP; 48570 A; 35113 C; 34446 G; 47914 T; 0 U; 0 Other;

  Query Match             90.5%;  Score 455.2;  DB 12;  Length 166043;
  Best Local Similarity   96.8%;  Pred. No. 1.6e-114;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145789 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 145730

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||| |||||||     ||||||||||||
Db     145729 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 145677
```

```
Qy       121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145676 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 145617

Qy       181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
             |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db    145616 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 145557

Qy       241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
             |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    145556 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 145497

Qy       301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
             |||| |||||||||||||| |||||||||||||||||||||||||||| |||||||||
Db    145496 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 145437

Qy       360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    145436 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 145377

Qy       420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145376 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 145317

Qy       480 CCAAATATTTGGAAACTGATGTCT 503
             |||||||||||||||||||||||||
Db    145316 CCAAATATTTGGAAACTGATGTCT 145293
```

RESULT 6
ACC79611
ID ACC79611 standard; DNA; 1184 BP.
XX
AC ACC79611;
XX
DT 06-AUG-2003 (first entry)
XX
DE Mouse peroxisome proliferator activated protein-gamma-2 promoter SEQ:7.
XX
KW Mouse; Stat5; fat cell precursor; doxycycline; anorectic; obesity;
KW peroxisome proliferator activated protein-gamma-2; promoter; ds.
XX
OS Mus musculus.
XX
PN WO2003018803-A1.
XX
PD 06-MAR-2003.
XX
PF 22-FEB-2002; 2002WO-JP001605.
XX
PR 23-AUG-2001; 2001JP-00253574.
XX
PA (HELI-) HELIX RES INST INC.
XX
PI Sato H, Wakao R, Wakao H;
XX
DR WPI; 2003-290072/28.
XX
PT Preparation of established fat cell precursors for application in
PT screening candidate compounds for controlling induction of
PT differentiation into fat cells, useful in treating e.g. obesity.
XX
PS Example 5; Page 56; 66pp; Japanese.
XX
CC The present invention describes established fat cell precursors with a
CC DNA structure where the Stat5 gene is ligated in a state which allows its
CC expression downstream of a promoter domain which can be induced by
CC doxycycline. Also described: (1) screening compounds with an activity of
CC controlling the induction of differentiation into fat cells comprising:
CC (a) inducing the differentiation into fat cells in the fat cell
CC precursors; (b) contacting a test compound with these cells after
CC induction of their differentiation into fat cells; (c) measuring extent
CC of the differentiation into fat cells of such cells; and (d) comparing
CC the extent of differentiation into fat cells with that observed in a
CC control for evaluation of activity of the test compound, during which the
CC order of steps (a) and (b) can be changed or carried out simultaneously
CC before steps (c) and (d); (2) screened compounds inhibiting the
CC differentiation of fat cell precursors into fat cells; and (3) screened
CC compounds promoting the differentiation of fat cell precursors into fat
CC cells. Stat5 has anorectic activity. The method is useful for the
CC preparation of established fat cell precursors for application in
CC screening candidate compounds for controlling induction of
CC differentiation into fat cells, which are useful in treating e.g.
CC obesity. The present sequence represents the mouse peroxisome
CC proliferator activated protein-gamma-2 (PPAR-gamma) promoter, which is
CC used in the exemplification of the present invention
XX
SQ Sequence 1184 BP; 362 A; 227 C; 203 G; 392 T; 0 U; 0 Other;

  Query Match   43.0%; Score 216.4; DB 10; Length 1184;
  Best Local Similarity 77.6%; Pred. No. 2.2e-49;
  Matches 354; Conservative   0; Mismatches   86; Indels   16; Gaps   7;

```
Qy        57 AACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGA 116
             ||  |||  || || ||||| | | | | |||||||||||  |||||||||       ||
Db       677 AAGAATGTGTATATGTTTGAGTACAAGAATATTCTTCAGAT-GTGTGATTAG------GA 729

Qy       117 GATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTT-TCTGGTATTT 175
             | ||||||||  |||| | ||  ||  |||||||||| ||||||| | || |||||| |||
Db       730 GTTTCAACCAAAGATAAATACTTAAGAAAAACTTTGGCCAAATACGTTTATCTGGTGTTT 789

Qy       176 CATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCA---TTGTGAATTATACAACAAT 232
             |||||   ||||||| | |||||| |||| | ||||||    |||| ||| |||  ||||
Db       790 CATAACTTAGAGATTAAGGTTTTCTATTTTAAAAGCCACTGGTGTGTATTTTACTGCAAT 849

Qy       233 AAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTT 292
                 |||    | | |||||||||| ||||| ||||| |||||| || ||| ||||||||
Db       850 TTTAAAAAGCAATCAATATTGAACAATCTCTGCTCTGGTAATTCCAACTACTGTACAGTT 909

Qy       293 CACGCCCCTCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTAT 351
             |||||||||||   ||| ||||  ||| |||||| ||||||||| ||| ||||  || |
Db       910 CACGCCCCTCACAGAACAGTGAATGTGTGGGTCACTGGCGAGACAATGTAGCAACGTTTT 969

Qy       352 CCCTGTAATGTACCAAGTCTTGCCA---GAGCAGTGAACATTATGACACAACTTTTTGTC 408
             || |||||||||||||||||||||    |||||  | |||||||||||| | |||||||
Db       970 CCTTGTAATGTACCAAGTCTTGCCAAAGCAGCAGACAGCATTATGACACACCATTTTGTC 1029

Qy       409 ACAGCTGGCTCCTAAT-AGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTA 467
             ||| |||||||  | | |||||||||||||||||||| |||   | ||||||||||||||
Db      1030 ACAACTGGCTCTCAGTCAGGACAGTGCCAGCCAATTCAGGCCTGATTCTTTCTGTGTTTA 1089

Qy       468 TTCCCATCTCTCCCAAATATTTGGAAACTGATGTCT 503
             |||||| |||||||||||||||| |||||| |||||
Db      1090 TTCCCACCTCTCCCAAATATTTGAAAACTGGTGTCT 1125
```

RESULT 7
AAX19064
ID AAX19064 standard; DNA; 177 BP.
XX
AC AAX19064;
XX
DT 13-MAY-1999 (first entry)
XX
DE Human PPAR-gamma-2 proximal promoter.
XX
KW Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;
KW regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;
KW lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;
KW insulin resistance; non-insulin-dependent diabetes mellitus;
KW polycystic ovary syndrome; gastrointestinal tract; Crohn's disease;
KW inflammatory bowel disease; ulcerative colitis; bowel cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9905161-A1.

XX
PD 04-FEB-1999.
XX
PF 24-JUL-1998; 98WO-US015411.
XX
PR 25-JUL-1997; 97US-0053692P.
XX
PA (LIGA-) LIGAND PHARM INC.
PA (INSP) INST PASTEUR.
XX
PI Briggs MR, Saladin RS, Auwerx J, Fajas L;
XX
DR WPI; 1999-142844/12.
XX
PT Newly isolated nucleic acid comprising a control region of a human
PT peroxisome proliferator activated receptor (PPAR) gamma gene - useful for
PT identifying modulators that are useful in treating diseases associated
PT with abnormal levels of human PPAR-gamma gene expression.
XX
PS Disclosure; Page 87; 102pp; English.
XX
CC The present invention describes an isolated, purified or enriched nucleic
CC acid comprising a control region of a human peroxisome proliferator
CC activated receptor gamma (PPAR-gamma) gene. The nucleic acids are useful
CC for screening for agents capable of modulating the expression of a human
CC PPAR-gamma gene. These agents (modulators) form pharmaceutical
CC compositions that are useful for treating diseases associated with
CC high/low levels of human PPAR-gamma gene expression. The diseases include
CC obesity, anorexia, cachexia, lipodystrophy, lipomas, liposarcomas,
CC abnormalities associated with anti-human immunodeficiency virus (HIV)
CC treatment, insulin resistance, non-insulin-dependent diabetes mellitus
CC (NIDDM), polycystic ovary syndrome, diseases of the gastrointestinal (GI)
CC tract, inflammatory bowel disease, Crohn's disease, ulcerative colitis
CC and bowel cancer. The nucleic acids are useful for studying the role of
CC the PPAR-gamma gene in various diseases and disorders. The structure of
CC PPAR-gamma enables genetic studies of PPAR- gamma mutations in humans,
CC and evaluation of its role in disorders like insulin resistance, NIDDM,
CC and diseases associated with altered adipose tissue function, like
CC obesity and lipodystrophic syndromes. The nucleic acids are also useful
CC for gene therapy and the production of transgenic animals, which are
CC useful in screening assays. The control regions of the nucleic acids
CC enable screening for modulators of the human PPAR-gamma gene, which are
CC useful in designing drugs for treating disorders or diseases associated
CC with the level of PPAR-gamma gene expression. The present sequence
CC represents the human PPAR-gamma-2 proximal promoter
XX
SQ Sequence 177 BP; 44 A; 35 C; 29 G; 69 T; 0 U; 0 Other;

Query Match 10.3%; Score 52; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 452 GTCCTTTCTGTGTTTATTCCCATCTCTCCCAAATATTTGGAAACTGATGTCT 503
||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 GTCCTTTCTGTGTTTATTCCCATCTCTCCCAAATATTTGGAAACTGATGTCT 52

RESULT 8
ADO47191
ID ADO47191 standard; DNA; 200000 BP.
XX
AC ADO47191;
XX
DT 15-JUL-2004 (first entry)
XX
DE DNA sequence of a human immunoglobulin heavy chain variable region.
XX
KW immunoglobulin; B cell; germline; region V; region D; region J;
KW recombination-promoting protein; recombination activating gene 1; RAG-1;
KW RAG-2; ss; terminal deoxynucleotidyltransferase; TdT; V(D)J recombinase.
XX
OS Homo sapiens.
XX
PN WO2004029249-A1.
XX
PD 08-APR-2004.
XX
PF 30-SEP-2003; 2003WO-AU001286.
XX
PR 30-SEP-2002; 2002US-0415024P.
XX
PA (CENT-) CENTENARY INST CANCER MEDICINE & CELL BI.
XX
PI Jolly C;
XX
DR WPI; 2004-316126/29.
XX
PT New vector comprising one or more immunoglobulin regions selected from V,
PT D, and J regions, a 5' flanking region, and a 3' flanking region, useful
PT for reverting cell lines to a pro-B cell-like state or to a germline-like
PT state.
XX
PS Example 1; Page 136-185; 382pp; English.
XX
CC The specification describes a method for the preparation of
CC immunoglobulins. The method uses a vector for reverting cell lines to a
CC pro-B cell-like state or to a germline-like state, by replacing one or
CC more of the immunoglobulin regions V, D, and J of the rearranged
CC immunoglobulin gene with the germ-line or unrearranged V, D or J regions.
CC The vector can comprise a polynucleotide sequence encoding a
CC recombination-promoting protein, selected from recombination activating
CC gene 1 (RAG-1), RAG-2, terminal deoxynucleotidyltransferase (TdT). These
CC proteins collectively constitute a V(D)J recombinase. The method is
CC useful for the preparation of immunoglobulin heavy and/or light chains
CC and their fragments. The present sequence encodes a human immunoglobulin
CC heavy chain variable region, and is used as a template to construct
CC vectors for use in the method of the invention.
XX
SQ Sequence 200000 BP; 59904 A; 41194 C; 41535 G; 57367 T; 0 U; 0 Other;

Query Match 9.0%; Score 45.2; DB 12; Length 200000;
Best Local Similarity 48.1%; Pred. No. 0.19;
Matches 128; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

Qy 30 TCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTT 89
Db 139429 TGAGAACAATGGAAAGCTTTTCTTCTAAGATAAGGAATAAGAAAAGAATGCCCAGTCTTA 139488

Qy 90 CTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACT 149
Db 139489 ATAATTCTAATCAGCATATTGATGATAGTTTTAACCATAGAAATTTGAGAAAGAAAAAAA 139548

Qy 150 TTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAA 209
Db 139549 TAGTAAAAGGCATACAAATTGAGGGGAATAAGTTAAATTGTTTCTGTTTACAGACAATAT 139608

Qy 210 GCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTG 269
Db 139609 AACTTTATAAATTCCAAAAAAATAAAAAATTCTGAAAAAACAGCCACTGAAAACTAACAA 139668

Qy 270 ATAATTCTAAATACAGTACAGTTCAC 295
Db 139669 ACAAATTTAGTAACATTGCAGTATAC 139694

RESULT 9
ABQ72637/c
ID ABQ72637 standard; cDNA; 3162 BP.
XX
AC ABQ72637;
XX
DT 03-SEP-2002 (first entry)
XX
DE Human MDDT encoding cDNA SEQ ID NO 189.
XX
KW Human; MDDT; disease detection and treatment molecule polynucleotide;

KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;
KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antigout;
KW neuroprotective; antirheumatic; antiarthritic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200240715-A2.
XX
PD 23-MAY-2002.
XX
PF 06-SEP-2001; 2001WO-US027628.
XX
PR 05-SEP-2000; 2000US-0229747P.
PR 05-SEP-2000; 2000US-0229748P.
PR 05-SEP-2000; 2000US-0229749P.
PR 05-SEP-2000; 2000US-0229750P.
PR 05-SEP-2000; 2000US-0229751P.
PR 05-SEP-2000; 2000US-0230583P.
PR 06-SEP-2000; 2000US-0230505P.
PR 06-SEP-2000; 2000US-0230514P.
PR 06-SEP-2000; 2000US-0230515P.
PR 06-SEP-2000; 2000US-0230517P.
PR 06-SEP-2000; 2000US-0230518P.
PR 06-SEP-2000; 2000US-0230519P.
PR 06-SEP-2000; 2000US-0230595P.
PR 06-SEP-2000; 2000US-0230597P.
PR 06-SEP-2000; 2000US-0230598P.
PR 06-SEP-2000; 2000US-0230599P.
PR 06-SEP-2000; 2000US-0230610P.
PR 06-SEP-2000; 2000US-0230865P.
PR 06-SEP-2000; 2000US-0230988P.
PR 06-SEP-2000; 2000US-0230989P.
PR 07-SEP-2000; 2000US-0230951P.
PR 07-SEP-2000; 2000US-0231163P.
PR 07-SEP-2000; 2000US-0231167P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;
PI Hillman JL, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE;
PI Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B;
PI Roseberry AM, Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V;
PI Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
XX
DR WPI; 2002-527544/56.
DR P-PSDB; ABP51420.
XX
PT Novel human disease detection and treatment polypeptide, useful in
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.
PT AIDS.
XX
PS Claim 1; Page 407-408; 618pp; English.
XX
CC The invention relates to an isolated human disease detection and
CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a
CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
CC specification, a naturally occurring polypeptide comprising a sequence
CC having at least 90% identity to (I) or a biologically active or
CC immunogenic fragment of (I). (I) is useful for screening a compound for
CC effectiveness as an agonist or antagonist, for screening a compound that
CC specifically binds (I) or modulates the activity of (I), and for
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
CC screening a compound for effectiveness in altering expression of a target
CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
CC detecting MDDT in a sample or for assessing toxicity of a test compound,
CC in a diagnostic test for a condition or a disease associated with the
CC expression of MDDT in a biological sample, for detecting (I) in a sample,
CC and for purifying (I) from a sample. A composition comprising (I), an
CC agonist or antagonist is useful for treating a disease or condition
CC associated with decreased or increased expression of functional MDDT. (I)
CC or (II) are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of MDDT, where the disorders are
CC selected from a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, psoriasis, and cancer and an
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
CC rheumatoid arthritis. (II) are useful for creating knockin humanised
CC animals or transgenic animals to model human diseases, in somatic or
CC germline gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences
XX
SQ Sequence 3162 BP; 971 A; 580 C; 668 G; 943 T; 0 U; 0 Other;

Query Match 8.8%; Score 44.2; DB 6; Length 3162;
Best Local Similarity 46.3%; Pred. No. 0.092;
Matches 145; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

```
Qy         13 CTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATAT 72
              ||| | ||||||| || ||  ||   | ||      ||  |||||||   | || |  | |
Db       2842 CTTCGAAATATAAAATATCTATAAATTTCTATGTACATAGTTTAAAAATTTGAAAATTTT 2783

Qy         73 TTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATA 132
              |  ||     ||          | | || |  |  ||| ||       | |||  |   |
Db       2782 TCTAAAATATAAAAGAAAACTAAAAAAAGACAGAAGAAATGCTTCCCTTAACATGTTTTC 2723

Qy        133 GACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTA 192
              | |||| ||  ||| ||    |  |  |    |     | |   | ||| || |
Db       2722 CTATCAGAAATGAACTTTTTCCAGCTAATTCACTTCTATTAAGTAAACCAAGGAGCTTGA 2663

Qy        193 AGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATT 252
              | || |  | || | ||   ||    |||  ||  |||||  ||||  ||  ||| | ||
Db       2662 AATTCTAAAGTTTAAAAAGGATCAAAAATAGTAATACAATCTAAAAATCATATGGTTCTT 2603

Qy        253 GAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCCTCACGAGACACT 312
              |          || | |  |||  | |     |  ||| | | |   ||   |  | ||
Db       2602 GGCTTCAGAATTTTATTTTAAAGGTTATACAGGATCAGATTATGAGACTGTTGCAAACCT 2543

Qy        313 GAACATGTGGTCA 325
              |  |   ||||
Db       2542 GGGTAATCTGTCA 2530
```

RESULT 10
ABL33191/c
ID ABL33191 standard; DNA; 6215 BP.
XX
AC ABL33191;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1164.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.

```
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1164; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6215 BP; 1813 A; 169 C; 1450 G; 2783 T; 0 U; 0 Other;

  Query Match             8.7%;  Score 43.6;  DB 6;  Length 6215;
  Best Local Similarity   58.5%;  Pred. No. 0.17;
  Matches   76;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy        167 CTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATAC 226
              ||   ||||| ||| ||| | |   ||  || |||||||| ||   | | | ||| |
Db       5118 CTTTCATTTCTTAACCAATAAACCAAACATTACCATTTAAAAATTTAATATTAATAAACT 5059

Qy        227 AACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGT 286
              || ||   ||||   || |  | |     |  || ||||||| ||||  |  | ||||
Db       5058 AAAAACTCAAAAAAAAAATTAACAACATTCCTTCCCTTCTCTTATAACACATACTACAAA 4999

Qy        287 ACAGTTCACG 296
              |   ||||||
Db       4998 ATTATTCACG 4989


RESULT 11
ADI03931/c
ID   ADI03931 standard; DNA; 99916 BP.
XX
AC   ADI03931;
XX
DT   22-APR-2004  (first entry)
XX
DE   Human enzyme protein encoding genomic DNA.
XX
KW   Enzyme protein; drug screening; disease diagnosis; human; gene therapy;
KW   chromosome 6; gene; ds; glucuronyltransferase.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2614. .96915
FT                   /*tag=  b
FT                   /product= "enzyme protein"
FT                   /note= "contains introns"
FT   exon            2614. .3204
FT                   /*tag=  a
FT                   /number= 1
FT   intron          3205. .64312
FT                   /*tag=  c
FT                   /number= 1
FT   exon            64313. .64457
FT                   /*tag=  d
FT                   /number= 2
FT   intron          64458. .96602
FT                   /*tag=  e
FT                   /number= 2
FT   exon            96603. .96751
FT                   /*tag=  f
FT                   /number= 3
FT   intron          96752. .96831
FT                   /*tag=  g
FT                   /number= 3
FT   exon            96832. .96915
FT                   /*tag=  h
FT                   /number= 4
XX
PN   WO200268657-A2.
XX
PD   06-SEP-2002.
XX
PF   08-FEB-2002; 2002WO-US003623.
XX
PR   26-FEB-2001; 2001US-0270871P.
PR   26-MAR-2001; 2001US-00816095.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Gan W,  Yan C,  Merkulov GV,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2002-713380/77.
DR   P-PSDB; ADI03930.
XX
PT   New human enzyme proteins, useful for treating or diagnosing disorders
PT   associated with abnormal expression of the protein, in drug screening
PT   assays and pharmacogenomic analysis.
XX
PS   Claim 4; SEQ ID NO 3; 127pp; English.
XX
CC   The invention relates to a novel isolated enzyme protein and encoding
CC   polynucleotides. The protein shows a high degree of similarity to a
CC   glucuronyltransferase cloned from a rabbit brain cDNA library. The
CC   peptides and nucleic acid molecules are useful as models for the
CC   development of human therapeutic targets, aid in the identification of
CC   therapeutic proteins, and serve as targets for the development of human
CC   therapeutic agents. The peptide may be used in drug screening assays, in
CC   assays to determine the biological activity of the protein, to raise
CC   antibodies or to elicit another immune response, as a reagent in assays
CC   designed to quantitatively determine levels of the protein in biological
CC   fluids, or as markers for tissues in which the corresponding protein is
CC   preferentially expressed. The human enzyme protein is also useful for
CC   diagnosing a disease, predisposition to a disease, or treating a disorder
CC   characterized by an absence of, inappropriate or unwanted expression of
CC   the protein. The antibodies are useful in pharmacogenomic analysis, for
CC   inhibiting protein function, or for tissue typing. The nucleic acid
CC   molecules are useful as probes, primers, chemical intermediates, or in
CC   biological assays. The present sequence represents a human enzyme protein
CC   encoding genomic DNA.
XX
SQ   Sequence 99916 BP; 27736 A; 18701 C; 19334 G; 29032 T; 0 U; 5113 Other;

  Query Match             8.3%;  Score 41.8;  DB 6;  Length 99916;
  Best Local Similarity   50.8%;  Pred. No. 1.3;
  Matches  100;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy         42 TGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATG 101
              ||  || ||| |  | | | || | |  ||    ||  |  ||  | |  | | | |
Db       8890 TGTAACTAACCTGCACATTGTGCACATGTACCCTAAAACTTAAAGTATAATAATAATAAA 8831

Qy        102 TGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAA 161
                | | | | || |||   | ||||| | |   ||| | |||| |       | |||||
Db       8830 ATAAAATAAAATAAAGAAATATCCAGGAAGATCAACCCACAGAACAAAGAAAGCTAATAA 8771

Qy        162 GCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAAT 221
                |   | | || || ||  || | |||| | ||||| || ||| |    |     ||
Db       8770 AGTACATTTTCTTGCAAAAAAAAAAAATTTTATTTTTCAATGTAATACATTAACAACAAA 8711
```

```
Qy        222 TATACAACAATAAAAAA 238
              |   || ||  || ||
Db       8710 AAGGAAAAAAGTAACAA 8694


RESULT 12
ADQ97587/c
ID   ADQ97587 standard; DNA; 116297 BP.
XX
AC   ADQ97587;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer associated sequence HD10-009, SEQ ID 564.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 564; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 116297 BP; 33950 A; 22909 C; 23278 G; 36160 T; 0 U; 0 Other;

  Query Match             8.3%;  Score 41.8;  DB 12;  Length 116297;
  Best Local Similarity   50.8%;  Pred. No. 1.4;
  Matches  100;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy         42 TGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATG 101
              ||  || ||| |  | | | || | |  ||    ||  |  ||  | |  | | | |
Db      16889 TGTAACTAACCTGCACATTGTGCACATGTACCCTAAAACTTAAAGTATAATAATAATAAA 16830

Qy        102 TGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAA 161
              |  | | | || |||   | ||||| | |   ||| | |||| |       | |||||
Db      16829 ATAAAATAAAATAAAGAAATATCCAGGAAGATCAACCCACAGAACAAAGAAAGCTAATAA 16770

Qy        162 GCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAAT 221
              |   | | || || ||  || | |||| | ||||| || ||| |     |      ||
Db      16769 AGTACATTTTCTTGCAAAAAAAAAAAATTTTATTTTTCAATGTAATACATTAACAACAAA 16710

Qy        222 TATACAACAATAAAAAA 238
              |   || ||  || ||
Db      16709 AAGGAAAAAAGTAACAA 16693


RESULT 13
ABL56203
ID   ABL56203 standard; DNA; 32392 BP.
XX
AC   ABL56203;
XX
DT   01-JUL-2002  (first entry)
XX
DE   AmEPV genome fragment#5.
XX
KW   AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
KW   genetic deficiency disorder; ds.
XX
OS   Amsacta moorei entomopoxvirus.
XX
PN   WO200212526-A2.
XX
PD   14-FEB-2002.
XX
PF   10-AUG-2001; 2001WO-US025287.
XX
PR   10-AUG-2000; 2000US-0224479P.
PR   14-SEP-2000; 2000US-00662254.
XX
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Moyer RW,  Li Y,  Bawden AL;
XX
DR   WPI; 2002-227161/28.
XX
PT   Novel recombinant entomopox virus vector useful for delivering
PT   polynucleotide encoding protein to vertebrate cell, comprises
PT   polynucleotide encoding protein operably linked with heterologous
PT   promoter sequence.
XX
PS   Disclosure; Page 226-242; 326pp; English.
XX
CC   The invention relates to a recombinant entomopox virus (EPV) vector,
CC   comprising a polynucleotide encoding a protein operably linked with a
CC   heterologous promoter sequence. The invention also concerns methods for
CC   providing gene therapy for genetic deficiency disorders. Vectors of the
CC   invention are useful for delivering a polynucleotide encoding a protein
CC   to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC   The vector is introduced into the vertebrate cell by infection in a viral
CC   particle, or by transfection, transduction, or injection either in vitro
CC   or in vivo. The vector is useful for the delivery and expression of
CC   biologically useful proteins in gene therapy protocols, and for
CC   delivering large DNA segments for engineering of vertebrate cells.
CC   Polynucleotides of the invention have applications in techniques such as
CC   their use as insertion sites for foreign genes of interest, hybridisation
CC   probes, for chromosome and gene mapping, in PCR technologies, and in the
CC   production of sense or antisense nucleic acids. Vectors of the invention
CC   provide for stable integration and expression of heterologous DNA in host
CC   cells, and are adapted for accepting large heterologous polynucleotide
CC   inserts which can be delivered in an infected or transformed cell and
CC   expressed in a stable fraction. The current sequence represents a
CC   fragment of the genome of the genus B entomopoxvirus from amsacta moorei
CC   (AmEPV)
XX
SQ   Sequence 32392 BP; 13748 A; 2577 C; 2550 G; 13517 T; 0 U; 0 Other;

  Query Match             8.2%;  Score 41.2;  DB 6;  Length 32392;
  Best Local Similarity   48.7%;  Pred. No. 1.3;
  Matches  112;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy         14 TTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATT 73
              |||   ||||||| |   |  |  ||  | |||  |    |   || |||||| |  |||
Db       4783 TTATACATATAACTTAATATAATTATATTAATATTATTATAATAAATATGGAAAATTATC 4842

Qy         74 TGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAG 133
              | |   | |||      | | |  |  | | | |||||| | |  | | |  |||
Db       4843 ATATTATTATATTAACAATTAAAAGAAATTCTGACAGATTACAAAAACTAGAAAATATAT 4902

Qy        134 ACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAA 193
              |      |||  | |     |||||   || | || | |  || | || |||||  |
Db       4903 TATCTTGTCAAAATTTATTATATAATAAAGATTATAGTGTATTTTATGGAATAGATTATA 4962
```

```
Qy        194 GTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAA 243
Db       4963 AAAATATAAATAAAAATAATTTAAAAAATATATGTAAAAAAGGATTTAAA 5012
```

RESULT 14
ADJ40655/c
ID ADJ40655 standard; cDNA; 2000 BP.
XX
AC ADJ40655;
XX
DT 06-MAY-2004 (first entry)
XX
DE Plant cDNA #1655.
XX
KW Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW antifungal.
XX
OS Eukaryota.
XX
PN US2004016025-A1.
XX
PD 22-JAN-2004.
XX
PF 26-SEP-2002; 2002US-00260238.
XX
PR 26-SEP-2001; 2001US-0325277P.
PR 26-SEP-2001; 2001US-0325448P.
PR 04-APR-2002; 2002US-0370620P.
XX
PA (BUDW/) BUDWORTH P.
PA (MOUG/) MOUGHAMER T.
PA (BRIG/) BRIGGS S P.
PA (COOP/) COOPER B.
PA (GLAZ/) GLAZEBROOK J.
PA (GOFF/) GOFF S A.
PA (KATA/) KATAGIRI F.
PA (KREP/) KREPS J.
PA (PROV/) PROVART N.
PA (RICK/) RICKE D.
PA (ZHUT/) ZHU T.
XX
PI Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
PI Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX
DR WPI; 2004-190374/18.
XX
PT New rice promoter, useful for manipulating crop plants to alter or
PT improve phenotypic characteristics, e.g. produce large quantities of oil
PT or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT or high nutritional value.
XX
PS Claim 26; SEQ ID NO 1655; 230pp; English.
XX
CC The invention relates to plant nucleotide sequences that direct seed-,
CC leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC or constitutive transcription of an operatively linked nucleic acid
CC segment. The invention also relates to a method for augmenting a plant
CC genome and a method of identifying a gene, where its expression is
CC altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC encode are useful for manipulating crop plants to alter or improve
CC phenotypic characteristics, to produce large quantities of oil or
CC proteins, to incur resistance to insecticides, viruses or fungi, and to
CC incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC have a high nutritional value with reduced apical dominance or dwarfism,
CC early flowering or altered metabolic pathways. This sequence represents a
CC plant nucleic acid of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification but was obtained in
CC electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 2000 BP; 540 A; 381 C; 366 G; 713 T; 0 U; 0 Other;

Query Match 8.1%; Score 40.8; DB 12; Length 2000;
Best Local Similarity 48.0%; Pred. No. 0.69;
Matches 117; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```
Qy          4 AGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATG 63
Db       1937 ATTTCGTATGATAAACATAGAAAATAAATATAAGATGAAGATGTGTTCTTTCAACTAAAT 1878

Qy         64 GATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAA 123
Db       1877 GACACGATTTTGTATGCTTCATATAATTGCATATTTTATAATCATCGAGAAAACTATAAG 1818

Qy        124 CCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCA 183
Db       1817 CAATTGAGATGATTTGCTTTATGATATTTTAGATCAAATATTGCTTGATCGTCTAAAGGA 1758

Qy        184 AGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAA 243
Db       1757 AAAAACTAAGATGATGGGACACAAAGGACAATGTGGACCATAGAGCTATAAATAATTAAA 1698

Qy        244 GTGG 247
Db       1697 ATGG 1694
```

RESULT 15
ADQ97818/c
ID ADQ97818 standard; DNA; 109661 BP.
XX
AC ADQ97818;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human cancer associated sequence HD11-002, SEQ ID 795.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS Homo sapiens.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 795; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic formate directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 109661 BP; 30680 A; 19371 C; 19986 G; 35350 T; 0 U; 4274 Other;

Query Match 8.1%; Score 40.8; DB 12; Length 109661;
Best Local Similarity 50.5%; Pred. No. 2.5;
Matches 99; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```
Qy         76 AATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGAC 135
              | |||||| | | | |  | |  || ||  || |       |   |  |  |   |||
Db      15320 ACTTCCAAGTATGTTTATAAAATATATGTCTAAAACCATTAAGGTTGAGAAAATATACTG 15261

Qy        136 ACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGT 195
               |||| |||  |||    |||| ||  | |||  | | |||||  || || ||||| |
Db      15260 CATGAAAAAAATTTTTAAGTAAATTAGAATACTGCAAATGCATAATAAACAGCTTTAAAT 15201

Qy        196 TTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAA 255
                | |     |  |||    | ||  ||  || |||   ||  | |  || ||||
Db      15200 AGCTACCTGAGATGTCACTTGATGTTGATCAAAATATGAAAGGAGCTGGAGTTTAGTGAA 15141

Qy        256 CAGTCTCTTCTCTGAT 271
              ||   ||| |  || |
Db      15140 CACAGTCTCCAGTGTT 15125
```

Search completed: November 9, 2005, 00:12:39
Job time : 411.412 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 15:16:42 ; Search time 128.167 Seconds
(without alignments)
6421.701 Million cell updates/sec

Title: US-09-463-542-3_COPY_818_1320
Perfect score: 503
Sequence: 1 tagagtaagtaccttaggaa.....atatttggaaactgatgtct 503

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	455.2	90.5	1100	3	US-09-484-345-10	Sequence 10, Appl
	2	455.2	90.5	86877	4	US-09-949-016-15491	Sequence 15491, A
	3	455.2	90.5	86877	4	US-09-949-016-15492	Sequence 15492, A
	4	142.4	28.3	601	4	US-09-949-016-133111	Sequence 133111,
	5	142.4	28.3	601	4	US-09-949-016-133194	Sequence 133194,
c	6	41.8	8.3	99916	4	US-09-816-095-3	Sequence 3, Appli
c	7	40.4	8.0	1141	4	US-09-806-708B-22	Sequence 22, Appl
c	8	39.4	7.8	1055	4	US-09-806-708B-23	Sequence 23, Appl
c	9	38.8	7.7	53806	4	US-09-949-016-12572	Sequence 12572, A
	10	38.4	7.6	108310	4	US-09-949-016-16366	Sequence 16366, A
c	11	38.2	7.6	136264	4	US-09-949-016-12756	Sequence 12756, A
c	12	38.2	7.6	136265	4	US-09-949-016-13001	Sequence 13001, A
	13	38	7.6	1317	4	US-09-248-796A-1010	Sequence 1010, Ap
c	14	37.6	7.5	640681	4	US-09-790-988-1	Sequence 1, Appli
c	15	37	7.4	192506	4	US-09-949-016-15830	Sequence 15830, A
c	16	36.8	7.3	117838	4	US-09-949-016-17595	Sequence 17595, A
c	17	36.6	7.3	1806	4	US-09-746-359A-36	Sequence 36, Appl
	18	36.2	7.2	120609	4	US-09-949-016-13915	Sequence 13915, A
	19	35.6	7.1	601	4	US-09-949-016-155222	Sequence 155222,
c	20	35.4	7.0	147321	4	US-09-949-016-15450	Sequence 15450, A
	21	35.2	7.0	2193	4	US-09-011-143-5	Sequence 5, Appli
	22	35.2	7.0	2193	4	US-09-302-495-5	Sequence 5, Appli
	23	35.2	7.0	2193	4	US-10-079-616-5	Sequence 5, Appli
c	24	35.2	7.0	19124	2	US-08-487-826B-13	Sequence 13, Appl
	25	35.2	7.0	64638	4	US-09-949-016-11767	Sequence 11767, A
	26	35.2	7.0	64639	4	US-09-949-016-13520	Sequence 13520, A
c	27	35.2	7.0	251769	4	US-09-949-016-13185	Sequence 13185, A
c	28	35.2	7.0	251769	4	US-09-949-016-13186	Sequence 13186, A
c	29	35.2	7.0	266748	4	US-09-949-016-13187	Sequence 13187, A
c	30	35.2	7.0	266748	4	US-09-949-016-13188	Sequence 13188, A
	31	35	7.0	798	4	US-09-270-767-5570	Sequence 5570, Ap
	32	35	7.0	798	4	US-09-270-767-20852	Sequence 20852, A
c	33	35	7.0	966	4	US-09-270-767-6622	Sequence 6622, Ap
c	34	35	7.0	966	4	US-09-270-767-21904	Sequence 21904, A
	35	35	7.0	1457	4	US-09-270-767-436	Sequence 436, App
	36	35	7.0	1457	4	US-09-270-767-15718	Sequence 15718, A
	37	34.6	6.9	15861	4	US-09-949-016-17604	Sequence 17604, A
c	38	34.6	6.9	387902	4	US-09-949-016-14543	Sequence 14543, A
c	39	34.6	6.9	421883	4	US-09-949-016-12557	Sequence 12557, A
	40	34.4	6.8	601	4	US-09-949-016-155221	Sequence 155221,
	41	34.4	6.8	1141	4	US-09-806-708B-22	Sequence 22, Appl
	42	34.4	6.8	1458	4	US-09-248-796A-1619	Sequence 1619, Ap
c	43	34.4	6.8	43981	4	US-09-949-016-16294	Sequence 16294, A
	44	34.4	6.8	96987	4	US-09-949-016-14429	Sequence 14429, A
c	45	34.4	6.8	175265	4	US-09-949-016-16089	Sequence 16089, A

ALIGNMENTS

RESULT 1
US-09-484-345-10
; Sequence 10, Application US/09484345
; Patent No. 6159734
; GENERAL INFORMATION:
; APPLICANT: Robert McKay
; APPLICANT: Alexander H. Borchers
; APPLICANT: Brenda F. Baker
; TITLE OF INVENTION: ANTISENSE MODULATION OF PEROXISOME PROLIFERATOR-ACTIVATED RECEPTO
; FILE REFERENCE: RTS-0104
; CURRENT APPLICATION NUMBER: US/09/484,345
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 10
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (859)...(940)
US-09-484-345-10

Query Match 90.5%; Score 455.2; DB 3; Length 1100;
Best Local Similarity 96.8%; Pred. No. 1.8e-124;
Matches 488; Conservative 0; Mismatches 8; Indels 8; Gaps 2;

```
Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 298

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||| |||||||           ||||||||||||
Db        299 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 351

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 411

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db        412 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 471

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        472 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 531

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| |||||||||||||| |||||||||||||||||||||||||||| |||||||||
Db        532 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 591
```

```
Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        592 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 651

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 711

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db        712 CCAAATATTTGGAAACTGATGTCT 735


RESULT 2
US-09-949-016-15491
; Sequence 15491, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15491
;   LENGTH: 86877
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15491

  Query Match             90.5%;  Score 455.2;  DB 4;  Length 86877;
  Best Local Similarity   96.8%;  Pred. No. 1.1e-123;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1472 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 1531

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||||| |||||||        ||||||||||
Db       1532 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 1584

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1585 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 1644

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db       1645 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 1704

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1705 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 1764

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| ||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db       1765 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 1824

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db       1825 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 1884

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1885 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 1944

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db       1945 CCAAATATTTGGAAACTGATGTCT 1968


RESULT 3
US-09-949-016-15492
; Sequence 15492, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15492
;   LENGTH: 86877
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15492

  Query Match             90.5%;  Score 455.2;  DB 4;  Length 86877;
  Best Local Similarity   96.8%;  Pred. No. 1.1e-123;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1472 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 1531

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
              |||||||||||||||||||||||||||||||||| |||||||        ||||||||||
Db       1532 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 1584

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1585 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 1644

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
              |||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db       1645 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 1704

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       1705 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 1764

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
              |||| ||||||||||||||| |||||||||||||||||||||||||||||| |||||||||
Db       1765 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 1824

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db       1825 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 1884

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1885 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 1944

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db       1945 CCAAATATTTGGAAACTGATGTCT 1968


RESULT 4
```

```
US-09-949-016-133111
; Sequence 133111, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 133111
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-133111

  Query Match             28.3%;  Score 142.4;  DB 4;  Length 601;
  Best Local Similarity   99.3%;  Pred. No. 3.5e-32;
  Matches  143;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db          1 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 60

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 120

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db        121 CCAAATATTTGGAAACTGATGTCT 144


RESULT 5
US-09-949-016-133194
; Sequence 133194, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 133194
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-133194

  Query Match             28.3%;  Score 142.4;  DB 4;  Length 601;
  Best Local Similarity   99.3%;  Pred. No. 3.5e-32;
  Matches  143;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db          1 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 60

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 120

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
              |||||||||||||||||||||||||
Db        121 CCAAATATTTGGAAACTGATGTCT 144


RESULT 6
US-09-816-095-3/c
; Sequence 3, Application US/09816095
; Patent No. 6664084
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001147
;  CURRENT APPLICATION NUMBER: US/09/816,095
;  CURRENT FILING DATE:  2001-03-26
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 99916
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(99916)
;   OTHER INFORMATION: n = A,T,C or G
US-09-816-095-3

  Query Match              8.3%;  Score 41.8;  DB 4;  Length 99916;
  Best Local Similarity   50.8%;  Pred. No. 0.15;
  Matches  100;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy         42 TGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATG 101
              ||  || ||| |  | | | || | |  ||    ||  |  ||  | |  | | |  |
Db       8890 TGTAACTAACCTGCACATTGTGCACATGTACCCTAAAACTTAAAGTATAATAATAATAAA 8831

Qy        102 TGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAA 161
              |  | | | || |||   | ||||| | |   ||| | |||| |       | |||||
Db       8830 ATAAAATAAAATAAAGAAATATCCAGGAAGATCAACCCACAGAACAAAGAAAGCTAATAA 8771

Qy        162 GCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAAT 221
              |   | | || || ||  || | |||| | ||||| || ||| |    |      ||
Db       8770 AGTACATTTTCTTGCAAAAAAAAAAAATTTTATTTTTCAATGTAATACATTAACAACAAA 8711

Qy        222 TATACAACAATAAAAAA 238
              |   || ||  || ||
Db       8710 AAGGAAAAAAGTAACAA 8694


RESULT 7
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
```

```
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             8.0%;  Score 40.4;  DB 4;  Length 1141;
  Best Local Similarity   9.0%;  Pred. No. 0.062;
  Matches   37;  Conservative  175;  Mismatches  197;  Indels    0;  Gaps    0;

Qy         52 GTTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGA 111
              ::||   :::: : :: ::| : : ::   : :  |: ::: :  :::     :: :   :
Db        610 RDTTCTYVDVWADSWVWWYANWMRCRDVTYTRNNTYCKSYAHSYWYWSNNAMWYRRYSAR 551

Qy        112 TTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGT 171
               : :: ::| :   : :|  : : |   ::: :    |:|::: ::  :|::|||::::
Db        550 NWSSMARWTTRNNWWMSGBVRMRWAGTMWWRHWNNNNTDTRYYWWWKRWARBTTTVYDSM 491

Qy        172 ATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAA 231
                :::::   :: ::::: |  :   : ::|:::|   : :   ::::  |:  :|:
Db        490 CNAKSMWRGNNWRAMKMWWAANNDAGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYCC 431

Qy        232 TAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGT 291
                 :| : :  :  :: :: :: |      :  :: : |:  : :: |  :  :  ::
Db        430 NNNNNRACVWHKHKMWRWTWKYMWKAACNNNNBKAMYMRVAWMMYSRDTTNTDWMMWTSD 371

Qy        292 TCACGCCCCTCACGAGACACTGAACATGTGGTCACCGGCGAGACAGTGTGGCAATATTAT 351
               :::: : : : :::|        :   :| :::: :: :      |  | |:| ::  :
Db        370 WBWHWYTVDYTMMRAWNNNNNNNNWRBCKTTSWMWWMDHMNTHCTYGNNTWGSAYBMAAMS 311

Qy        352 CCCTGTAATGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACA 411
               :   |   ::|    ::  : |  :       :|::  | |:|    :::::| ::   :
Db        310 MWAAGASNBVTYNWCWRMTYMGKTMTNNNNNNKAWYYRTKTVAWCNNRYYYDTAVWTBKR 251

Qy        412 GCTGGCTCCTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCT 460
               ::  : :::  ::    : ::  :: ::: : :: : |: :::  :::
Db        250 NYKYCYAYBWYYBMYMGKHHWBWWRRABHRSWNMWWVKCRNKYMVSWHY 202


RESULT 8
US-09-806-708B-23/c
; Sequence 23, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
;   LENGTH: 1055
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1055)
;   OTHER INFORMATION: consensus sequence of A.t. and L.a. FAE1 promoters
US-09-806-708B-23

  Query Match             7.8%;  Score 39.4;  DB 4;  Length 1055;
  Best Local Similarity   21.5%;  Pred. No. 0.12;
  Matches   55;  Conservative   77;  Mismatches  124;  Indels    0;  Gaps    0;

Qy         38 ATGCTGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGAT 97
              ||  ||: | :|  ::| :  ||  ::: : :  :|  ::    :  | |:   | :: |
Db        565 ATTATGRGATSAKKRKTAWTTCTYRRRWAWGWRTWTANWMRCAWRTYTRTYCKSTAMSCT 506

Qy         98 AATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAA 157
               :    :: |     | :: ::| |   | ||  : | |   :|| |    | |::: ::
Db        505 YWGNAMWYAAYSAGNTSSMARWTTANNTAAGGYMRAAWAGTMWAAMANNNNTRTRYYAWW 446

Qy        158 ATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGT 217
                 :|::||| :::    :::::   :: | :|| |  :   : ::|||||   : |   |
Db        445 TRWARYTTTRYRSACNAKSMWRGATWRAATMTTAANNKAGAMWMWTTAAGNNTMTAAAAT 386

Qy        218 GAATTATACAACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCT 277
               :::  : | | |     :|||   : ::   :|::: :  |     :|  : :: :| ::
Db        385 KMMAAWCARAYCCNNNWAAACMAMKWKMWAWTWKYAWGAACNNNNKTAMYCRRAWAMYSA 326

Qy        278 AAATACAGTACAGTTC 293
               :  |: |   :: :::
Db        325 WTTTWWAAATSWWKWY 310


RESULT 9
US-09-949-016-12572/c
; Sequence 12572, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12572
;   LENGTH: 53806
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12572

  Query Match             7.7%;  Score 38.8;  DB 4;  Length 53806;
  Best Local Similarity   47.9%;  Pred. No. 0.89;
  Matches  112;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy         19 AATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTTGAAT 78
              | ||||| | |  |  | |     |||  ||  ||| ||  ||  ||| |||  | ||
Db      15953 AGTATAATAATAAATAAATAAATAAATAAGAATATTTTAAAAATAAATAAATAAATAAAA 15894

Qy         79 TCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACC 138
              |||| | ||||   | ||| ||  | |  | |   |  ||| ||       | | |||
Db      15893 GCCAACTGTTTCCCTATATATTGGCAATGAACAAGTGGAATTTAAAATAAAAAACACATT 15834

Qy        139 GAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTT 198
               |   | || || |  | ||     | |||||       | ||| |  | ||
Db      15833 ACCATTTACATTAGCACTCAAAACTACTTAGGTATAAATATAACAAAATGTGTACAAGAG 15774

Qy        199 CCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATT 252
              | || |||| |    |    || | |   | || | ||  |    || | | | |
Db      15773 CTATATAAGGAAAACTGCAAAACTCTGACAAAAGATAACAAAGTAGAGTAAAAT 15720


RESULT 10
US-09-949-016-16366
; Sequence 16366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
```

```
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16366
;   LENGTH: 108310
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(108310)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16366

  Query Match             7.6%;  Score 38.4;  DB 4;  Length 108310;
  Best Local Similarity   47.2%;  Pred. No. 1.6;
  Matches  117;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy         42 TGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATG 101
              | |||    |  | || |  |   |||  ||  |  || |   |  || | |  |     |
Db      16475 TTATATATATTTATATATAAATTATAAATTATGTAAATTATATATTATATAAATTATAAA 16534

Qy        102 TGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAA 161
              |   || | |||| | ||   |        || | |   | | |||   |      || ||
Db      16535 TATATATAAATTATATATATTATATAAATTATAAATATATATAAATTATATATAAATTAT 16594

Qy        162 GCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAAT 221
                |       |  | |||    |  | ||.||| || |  |||||    |      | | |||
Db      16595 AAATATATAAAAATTATATAAATTATATATAAATTATAAATTTATATAATTTATATAAAT 16654

Qy        222 TATACAACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAAT 281
              |||| |  ||| | ||||  |  |  ||  |  | | | |  |  |||| | || ||
Db      16655 TATATATAAATTATAAATATATATAAATTATATATATTATATATAATTATAAATATATAT 16714

Qy        282 ACAGTACA 289
              | | || |
Db      16715 AAATTATA 16722


RESULT 11
US-09-949-016-12756/c
; Sequence 12756, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12756
;   LENGTH: 136264
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(136264)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12756

  Query Match             7.6%;  Score 38.2;  DB 4;  Length 136264;
  Best Local Similarity   49.8%;  Pred. No. 2;
  Matches  126;  Conservative    0;  Mismatches  123;  Indels    4;  Gaps    1;

Qy         66 TACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACC 125
              || ||| |  |   |  | ||| |   | | ||||| |    |  | |||| | |||  |
Db      24356 TATATAATATATATTAATATTATATATTATATAATATATATTAATATTATATATTATA-- 24299

Qy        126 AGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAG 185
              |||  |  |   | |  | |  ||    ||   |  ||  |  |||  |  ||   ||
Db      24298 --TAATATATATTAATAATATATATTATATAATATATATTAATATTATATATTATATAAT 24241

Qy        186 AGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGT 245
              | || | | |  || | | |    ||     | |||   | | ||| | |||  |  |
Db      24240 ATATATTAATATTATATATTATATAATATATATTAATATTATATATTATATAATATATAT 24181

Qy        246 GGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCCTCACG 305
              |||||  | | | | | | | ||  | || |||    ||| |    ||  |||||
Db      24180 TAATATTATATATTATATAATATACTATATATATATATGAAACATTGGTGGCATCTCACA 24121

Qy        306 AGACACTGAACAT 318
              ||  |||  ||||
Db      24120 AGTAACTCCACAT 24108


RESULT 12
US-09-949-016-13001/c
; Sequence 13001, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13001
;   LENGTH: 136265
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(136265)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13001

  Query Match             7.6%;  Score 38.2;  DB 4;  Length 136265;
  Best Local Similarity   49.8%;  Pred. No. 2;
  Matches  126;  Conservative    0;  Mismatches  123;  Indels    4;  Gaps    1;

Qy         66 TACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACC 125
              || ||| |  |   |  | ||| |   | | ||||| |    |  | |||| | |||  |
Db      24356 TATATAATATATATTAATATTATATATTATATAATATATATTAATATTATATATTATA-- 24299

Qy        126 AGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAG 185
              |||  |  |   | |  | |  ||    ||   |  ||  |  |||  |  ||   ||
Db      24298 --TAATATATATTAATAATATATATTATATAATATATATTAATATTATATATTATATAAT 24241

Qy        186 AGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGT 245
              | || | | |  || | | |    ||     | |||   | | ||| | |||  |  |
Db      24240 ATATATTAATATTATATATTATATAATATATATTAATATTATATATTATATAATATATAT 24181

Qy        246 GGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCCTCACG 305
              |||||  | | | | | | | ||  | || |||    ||| |    ||  |||||
Db      24180 TAATATTATATATTATATAATATACTATATATATATATGAAACATTGGTGGCATCTCACA 24121

Qy        306 AGACACTGAACAT 318
```

```
              ||  |||  ||||
Db      24120 AGTAACTCCACAT 24108


RESULT 13
US-09-248-796A-1010
; Sequence 1010, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 60/074,725
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: US 60/096,409
;  PRIOR FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 1010
;   LENGTH: 1317
;   TYPE: DNA
;   ORGANISM: Candida albicans
US-09-248-796A-1010

  Query Match             7.6%;  Score 38;  DB 4;  Length 1317;
  Best Local Similarity   47.8%;  Pred. No. 0.34;
  Matches  110;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy         53 TTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGAT 112
              ||||||  ||  | |   ||| |||||    |  ||           |    |||| | |
Db        624 TTTAAATAATTTAAAACAATTGGAATTAATTAAATTAGGATCATCGAAAACATTAAAAAC 683

Qy        113 TAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTA 172
              |  ||||  ||      |   | | | ||| ||| |     |||||  ||  |||   | |
Db        684 TTTAGATATAAAATTAAAATTAGAACAAAAAAAATCCACTACCAAAGAAATTTTAATGAA 743

Qy        173 TTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAAT 232
              |  |  |  || |    | | |||    |  || || | | ||  | | | | |  |||
Db        744 ATGGAGTATTAATAATAATGATTTTAAATTACAAAAAACTAATGAAATTAAAAAATTAAT 803

Qy        233 AAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATA 282
              || ||     ||||  || | | |   | ||   | |  ||  |  |||
Db        804 TAATAACAATGGTGGTAATGGGAAATCATTTTTAATTAATATGGTTTATA 853


RESULT 14
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             7.5%;  Score 37.6;  DB 4;  Length 640681;
  Best Local Similarity   49.5%;  Pred. No. 5.6;
  Matches   97;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy         44 ATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTG 103
              |||   |  |||  | |||  |   | |  | || |    ||||| |    ||
Db     212263 ATAAATAAATTTCTAATATACAATAAAAAATAAAATATTTATTTTTTAACTATTGCACT 212204

Qy        104 ATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGC 163
              | ||  ||| |     | |   |     |  || | ||| ||   |||    ||     |
Db     212203 AATAATGATCATTATGTTATAAAAATTAATGCATCTAAAAAAGTTTTTAAAAAACAGTTC 212144

Qy        164 TTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTA 223
              ||| |  ||||| | ||     | || ||  ||||  || ||| ||   | | |   |||
Db     212143 TTTTTAATATTTTAAAAAATGAAAATATATATTTTTTATATAAAAATATAATTTCTTTTA 212084

Qy        224 TACAACAATAAAAAAT 239
              |||   | || |||||
Db     212083 TACCTAATTATAAAAT 212068


RESULT 15
US-09-949-016-15830/c
; Sequence 15830, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15830
;   LENGTH: 192506
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15830

  Query Match             7.4%;  Score 37;  DB 4;  Length 192506;
  Best Local Similarity   52.2%;  Pred. No. 5.1;
  Matches   82;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy         17 GGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTTGA 76
              |||| | | ||||| | |  |  |  |||   || |  | |  |||    ||  | |
Db     191028 GGAAGAGACCATTTAACTCTCCAACAAATAGTGACATCAAGAAAATGACACCAATTGTTG 190969

Qy         77 ATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACA 136
              |   |||||||||| |||| |  |       ||||| || | || ||| ||  || |||
Db     190968 AAAACAAATTTTTTTTCACAGTAAAATTGCAGAGAATACAAATACAAACACTGACAGATT 190909

Qy        137 CCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTAT 173
              |     || |  |  || ||  ||   || |   ||
Db     190908 ACTGCTTAAGATCTATCCTAACAAATTCTTTTCAAAT 190872


Search completed: November  8, 2005, 19:52:47
Job time : 131.167 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 18:06:53 ; Search time 650.857 Seconds
(without alignments)
6391.142 Million cell updates/sec

Title: US-09-463-542-3_COPY_818_1320
Perfect score: 503
Sequence: 1 tagagtaagtaccttaggaa.....atatttggaaactgatgtct 503

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	455.2	90.5	1100	24	US-10-181-176-10	Sequence 10, Appl
c	2	455.2	90.5	166043	19	US-10-235-192A-46	Sequence 46, Appl
	3	45.2	9.0	200000	20	US-10-672-764A-31	Sequence 31, Appl
c	4	44.2	8.8	3162	20	US-10-363-829-189	Sequence 189, App
c	5	43.6	8.7	6215	16	US-10-311-455-1164	Sequence 1164, Ap
c	6	41.8	8.3	99916	9	US-09-816-095-3	Sequence 3, Appli
c	7	41.8	8.3	99916	19	US-10-634-905-3	Sequence 3, Appli
	8	41.2	8.2	32392	22	US-10-706-635-27	Sequence 27, Appl
c	9	40.8	8.1	2000	18	US-10-260-238-1655	Sequence 1655, Ap
	10	40.8	8.1	59914	20	US-10-741-601-5619	Sequence 5619, Ap
c	11	40.8	8.1	101782	20	US-10-741-601-5661	Sequence 5661, Ap
c	12	40.4	8.0	1886	20	US-10-437-963-36653	Sequence 36653, A
	13	40.2	8.0	797	19	US-10-424-599-18984	Sequence 18984, A
c	14	40	8.0	8056	21	US-10-473-126-240	Sequence 240, App
c	15	39.8	7.9	350	20	US-10-674-124A-1055	Sequence 1055, Ap
c	16	39.4	7.8	350	18	US-10-242-535A-53661	Sequence 53661, A
c	17	39.4	7.8	350	19	US-10-085-783A-53661	Sequence 53661, A
	18	39.4	7.8	6191	16	US-10-311-455-1189	Sequence 1189, Ap
	19	39.2	7.8	629	13	US-09-925-065A-861470	Sequence 861470,
c	20	39.2	7.8	967	13	US-09-925-065A-37192	Sequence 37192, A
c	21	39.2	7.8	967	13	US-09-925-065A-37193	Sequence 37193, A
c	22	39.2	7.8	1943	13	US-09-925-065A-92097	Sequence 92097, A
c	23	39.2	7.8	6308	16	US-10-311-455-1443	Sequence 1443, Ap
c	24	39	7.8	6565	19	US-10-221-714A-188	Sequence 188, App
c	25	39	7.8	37515	20	US-10-433-793-28	Sequence 28, Appl
	26	38.8	7.7	505	14	US-10-027-632-248390	Sequence 248390,
	27	38.8	7.7	505	18	US-10-027-632-248390	Sequence 248390,
	28	38.8	7.7	1696	18	US-10-374-780A-1530	Sequence 1530, Ap
	29	38.8	7.7	1696	19	US-10-412-699B-1620	Sequence 1620, Ap
c	30	38.8	7.7	8056	21	US-10-473-126-386	Sequence 386, App
c	31	38.8	7.7	10945	16	US-10-240-453-228	Sequence 228, App
c	32	38.8	7.7	19634	21	US-10-473-126-156	Sequence 156, App
c	33	38.8	7.7	19634	21	US-10-473-126-302	Sequence 302, App
c	34	38.4	7.6	2258	26	US-11-097-143-29929	Sequence 29929, A
	35	38.2	7.6	586	14	US-10-027-632-197296	Sequence 197296,
	36	38.2	7.6	586	18	US-10-027-632-197296	Sequence 197296,
c	37	38.2	7.6	4152	22	US-10-706-635-9	Sequence 9, Appli
c	38	38.2	7.6	50000	22	US-10-706-635-25	Sequence 25, Appl
c	39	38.2	7.6	96589	19	US-10-052-482-214	Sequence 214, App
	40	38	7.6	629	13	US-09-925-065A-861469	Sequence 861469,
	41	38	7.6	629	13	US-09-925-065A-897143	Sequence 897143,
	42	38	7.6	2035	24	US-10-795-159-412	Sequence 412, App
c	43	38	7.6	3426	13	US-09-925-065A-676913	Sequence 676913,
c	44	38	7.6	3426	13	US-09-925-065A-676914	Sequence 676914,
c	45	38	7.6	908766	24	US-10-795-159-685	Sequence 685, App

ALIGNMENTS

RESULT 1
US-10-181-176-10
; Sequence 10, Application US/10181176
; Publication No. US20050186567A1
; GENERAL INFORMATION:
; APPLICANT: Robert McKay
; APPLICANT: Alexander H. Borchers
; APPLICANT: Brenda F. Baker
; TITLE OF INVENTION: ANTISENSE MODULATION OF PEROXISOME PROLIFERATOR-ACTIVATED RECEPTO
; TITLE OF INVENTION: GAMMA EXPRESSION
; FILE REFERENCE: RTSP-0328
; CURRENT APPLICATION NUMBER: US/10/181,176
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: PCT/US01/00952
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US09/484,345
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 10
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (859)...(940)
US-10-181-176-10

Query Match 90.5%; Score 455.2; DB 24; Length 1100;

```
  Best Local Similarity   96.8%;  Pred. No. 1.6e-108;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
Db        239 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 298

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
Db        299 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 351

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
Db        352 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 411

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
Db        412 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 471

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
Db        472 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 531

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
Db        532 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 591

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
Db        592 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 651

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
Db        652 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 711

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
Db        712 CCAAATATTTGGAAACTGATGTCT 735

RESULT 2
US-10-235-192A-46/c
; Sequence 46, Application US/10235192A
; Publication No. US20040043389A1
; GENERAL INFORMATION:
;  APPLICANT: McCarthy, Jeanette
;  TITLE OF INVENTION: Methods and Compositions for Identifying
;  TITLE OF INVENTION:  Risk Factors for Abnormal Lipid Levels and the Diseases
;  TITLE OF INVENTION:  and Disorders Associated Therewith
;  FILE REFERENCE: MMI-011
;  CURRENT APPLICATION NUMBER: US/10/235,192A
;  CURRENT FILING DATE:  2002-09-04
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46
;   LENGTH: 166043
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-235-192A-46

  Query Match             90.5%;  Score 455.2;  DB 19;  Length 166043;
  Best Local Similarity   96.8%;  Pred. No. 1.3e-107;
  Matches  488;  Conservative    0;  Mismatches    8;  Indels    8;  Gaps    2;

Qy          1 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 60
Db     145789 TAGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACT 145730

Qy         61 ATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATT 120
Db     145729 ATGGATACATATTTGAATTCCAAATTTTTCTTCAAATAATGT-------GATTAGAGATT 145677

Qy        121 CAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 180
Db     145676 CAACCAGGAATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAA 145617

Qy        181 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATG 240
Db     145616 GCAAGAGATTTAAGTTTTCCATTTAAGAAGCAATTGTGAATTTTACAACAATAAAAAATG 145557

Qy        241 CAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 300
Db     145556 CAAGTGGATATTGAACAGTCTCTGCTCTGATAATTCTAAATACAGTACAGTTCACGCCCC 145497

Qy        301 TCACGAGACACTGAACATGT-GGTCACCGGCGAGACAGTGTGGCAATATTATCCCTGTAA 359
Db     145496 TCACAAGACACTGAACATGTGGGTCACCGGCGAGACAGTGTGGCAATATTTTCCCTGTAA 145437

Qy        360 TGTACCAAGTCTTGCCAGAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 419
Db     145436 TGTACCAAGTCTTGCCAAAGCAGTGAACATTATGACACAACTTTTTGTCACAGCTGGCTC 145377

Qy        420 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 479
Db     145376 CTAATAGGACAGTGCCAGCCAATTCAAGCCCAGTCCTTTCTGTGTTTATTCCCATCTCTC 145317

Qy        480 CCAAATATTTGGAAACTGATGTCT 503
Db     145316 CCAAATATTTGGAAACTGATGTCT 145293

RESULT 3
US-10-672-764A-31
; Sequence 31, Application US/10672764A
; Publication No. US20040156832A1
; GENERAL INFORMATION:
;  APPLICANT: Jolly, Chris
;  TITLE OF INVENTION: Immunoglobulin Compositions and Methods
;  FILE REFERENCE: 13311.1001U
;  CURRENT APPLICATION NUMBER: US/10/672,764A
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 68
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 200000
;   TYPE: DNA
;   ORGANISM: Human
US-10-672-764A-31

  Query Match              9.0%;  Score 45.2;  DB 20;  Length 200000;
  Best Local Similarity   48.1%;  Pred. No. 2.6;
  Matches  128;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy         30 TCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTT 89
Db     139429 TGAGAACAATGGAAAGCTTTTCTTCTAAGATAAGGAATAAGAAAAGAATGCCCAGTCTTA 139488

Qy         90 CTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACT 149
Db     139489 ATAATTCTAATCAGCATATTGATGATAGTTTAACCATAGAAATTTGAGAAAGAAAAAAA 139548

Qy        150 TTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAA 209
Db     139549 TAGTAAAAGGCATACAAATTGAGGGGAATAAGTTAAATTGTTTCTGTTTACAGACAATAT 139608

Qy        210 GCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTG 269
Db     139609 AACTTTATAAATTCCAAAAAAATAAAAAATTCTGAAAAAACAGCCACTGAAAACTAACAA 139668

Qy        270 ATAATTCTAAATACAGTACAGTTCAC 295
Db     139669 ACAAATTTAGTAACATTGCAGTATAC 139694

RESULT 4
US-10-363-829-189/c
```

```
; Sequence 189, Application US/10363829
; Publication No. US20040142331A1
; GENERAL INFORMATION:
;  APPLICANT: Jackson, Stuart E.; Lincoln, Stephen E.;
;  APPLICANT:  Altus, Christina M.; Dufour, Gerard E.;
;  APPLICANT:  Chalup, Michael S.; Jackson, Jennifer L.;
;  APPLICANT:  Jones, Anissa L.; Yu, Jimmy Y.;
;  APPLICANT:  Wright, Rachel J.; Gietzen, Darryl;
;  APPLICANT:  Liu, Tommy F.; Yap, Pierre E.;
;  APPLICANT:  Dahl, Christopher R.; Momiyama, Monika G.;
;  APPLICANT:  Bradley, Diana L.; Rohatgi, Sameer D.;
;  APPLICANT:  Harris, Bernard; Roseberry Lincoln, Ann M.;
;  APPLICANT:  Gerstin,Jr., Edward H.; Peralta, Careyna H.;
;  APPLICANT:  David, Marie H.; Panzer, Scott R.;
;  APPLICANT:  Flores, Vincent Z.; Daffo, Abel;
;  APPLICANT:  Marwaha, Rakesh; Chen, Alice J.;
;  APPLICANT:  Chang, Simon C.; Au, Alan P.;
;  APPLICANT:  Inman, Rebekah R.
;  TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
;  FILE REFERENCE: PT-1183 USN
;  CURRENT APPLICATION NUMBER: US/10/363,829
;  CURRENT FILING DATE:  2003-03-05
;  PRIOR APPLICATION NUMBER: PCT/US01/27628
;  PRIOR FILING DATE: 2001-09-05
;  PRIOR APPLICATION NUMBER: US 60/229,751
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 60/229,749
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 60/229,750
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 60/229,747
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 60/229,748
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 60/230,583
;  PRIOR FILING DATE: 2000-09-05
;  PRIOR APPLICATION NUMBER: US 60/230,517
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: US 60/230,610
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: US 60/230,597
;  PRIOR FILING DATE: 2000-09-06
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 506
;  SOFTWARE: PERL Program
; SEQ ID NO 189
;   LENGTH: 3162
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: LI:203301.3:2000SEP08
US-10-363-829-189

  Query Match             8.8%;  Score 44.2;  DB 20;  Length 3162;
  Best Local Similarity   46.3%;  Pred. No. 0.79;
  Matches  145;  Conservative    0;  Mismatches  168;  Indels    0;  Gaps    0;

Qy         13 CTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATAT 72
              ||| | ||||||| || ||  ||   | ||      ||  |||||||   | || |  | |
Db       2842 CTTCGAAATATAAAATATCTATAAATTTCTATGTACATAGTTTAAAAATTTGAAAATTTT 2783

Qy         73 TTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATA 132
              |  ||      ||          | | ||| |  | ||| ||        | |||  |    |
Db       2782 TCTAAAATATAAAAGAAAACTAAAAAAAGACAGAAGAAATGCTTCCCTTAACATGTTTTC 2723

Qy        133 GACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTA 192
              | ||||  ||  ||| ||    |  |  |    |    | |  | ||| || |
Db       2722 CTATCAGAAATGAACTTTTTCCAGCTAATTCACTTCTATTAAGTAAACCAAGGAGCTTGA 2663

Qy        193 AGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATT 252
              | || |  | || | ||   ||    |||  ||  ||||| |||| ||  ||| | ||
Db       2662 AATTCTAAAGTTTAAAAAGGATCAAAAATAGTAATACAATCTAAAAATCATATGGTTCTT 2603

Qy        253 GAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCCTCACGAGACACT 312
              |          || | |  |||   | |     |  ||| | | |   ||   | |  ||
Db       2602 GGCTTCAGAATTTTATTTTAAAGGTTATACAGGATCAGATTATGAGACTGTTGCAAACCT 2543

Qy        313 GAACATGTGGTCA 325
              |   |   ||||
Db       2542 GGGTAATCTGTCA 2530


RESULT 5
US-10-311-455-1164/c
; Sequence 1164, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1164
;   LENGTH: 6215
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1164

  Query Match             8.7%;  Score 43.6;  DB 16;  Length 6215;
  Best Local Similarity   58.5%;  Pred. No. 1.5;
  Matches   76;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy        167 CTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATAC 226
              ||   ||||| ||| ||| | |   ||  || |||||||| ||   | | | ||| |
Db       5118 CTTTCATTTCTTAACCAATAAACCAAACATTACCATTTAAAAATTTAATATTAATAAACT 5059

Qy        227 AACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGT 286
              || ||   ||||   || |  | |      |  || ||||||| ||||  |  | ||||
Db       5058 AAAAACTCAAAAAAAAAATTAACAACATTCCTTCCCTTCTCTTATAACACATACTACAAA 4999

Qy        287 ACAGTTCACG 296
              |   ||||||
Db       4998 ATTATTCACG 4989


RESULT 6
US-09-816-095-3/c
; Sequence 3, Application US/09816095
; Patent No. US20020137164A1
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001147
;  CURRENT APPLICATION NUMBER: US/09/816,095
;  CURRENT FILING DATE:  2001-03-26
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 99916
;   TYPE: DNA
;   ORGANISM: Human
```

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(99916)
;   OTHER INFORMATION: n = A,T,C or G
US-09-816-095-3

  Query Match             8.3%;  Score 41.8;  DB 9;  Length 99916;
  Best Local Similarity   50.8%;  Pred. No. 15;
  Matches  100;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy         42 TGATACCAACGTTTAAACTATGGATACATATTGAATTCCAAATTTTTCTTCAGATAATG 101
              || || ||| |  | | | || | |  ||      ||  |  ||  | |  | | | |  |
Db       8890 TGTAACTAACCTGCACATTGTGCACATGTACCCTAAAACTTAAAGTATAATAATAATAAA 8831

Qy        102 TGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAA 161
                 |  |  | |  || |||    | |||||  | |    ||| |  ||||  |        | |||||
Db       8830 ATAAAATAAAATAAAGAAATATCCAGGAAGATCAACCCACAGAACAAAGAAAGCTAATAA 8771

Qy        162 GCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAAT 221
                 |    |  | || || ||  || |  |||| |  ||||| || ||| |      |      ||
Db       8770 AGTACATTTTCTTGCAAAAAAAAAAAAATTTTATTTTTCAATGTAATACATTAACAACAAA 8711

Qy        222 TATACAACAATAAAAAA 238
                |   || ||  || ||
Db       8710 AAGGAAAAAAGTAACAA 8694


RESULT 7
US-10-634-905-3/c
; Sequence 3, Application US/10634905
; Publication No. US20040067225A1
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu et al.
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001147-DIV
;  CURRENT APPLICATION NUMBER: US/10/634,905
;  CURRENT FILING DATE:  2003-08-06
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 99916
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(99916)
;   OTHER INFORMATION: n = A,T,C or G
US-10-634-905-3

  Query Match             8.3%;  Score 41.8;  DB 19;  Length 99916;
  Best Local Similarity   50.8%;  Pred. No. 15;
  Matches  100;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy         42 TGATACCAACGTTTAAACTATGGATACATATTGAATTCCAAATTTTTCTTCAGATAATG 101
              || || ||| |  | | | || | |  ||      ||  |  ||  | |  | | | |  |
Db       8890 TGTAACTAACCTGCACATTGTGCACATGTACCCTAAAACTTAAAGTATAATAATAATAAA 8831

Qy        102 TGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAA 161
                 |  |  | |  || |||    | |||||  | |    ||| |  ||||  |        | |||||
Db       8830 ATAAAATAAAATAAAGAAATATCCAGGAAGATCAACCCACAGAACAAAGAAAGCTAATAA 8771

Qy        162 GCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAAT 221
                 |    |  | || || ||  || |  |||| |  ||||| || ||| |      |      ||
Db       8770 AGTACATTTTCTTGCAAAAAAAAAAAAATTTTATTTTTCAATGTAATACATTAACAACAAA 8711

Qy        222 TATACAACAATAAAAAA 238
                |   || ||  || ||
Db       8710 AAGGAAAAAAGTAACAA 8694


RESULT 8
US-10-706-635-27
; Sequence 27, Application US/10706635
; Publication No. US20050014263A1
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XCZ1
;  CURRENT APPLICATION NUMBER: US/10/706,635
;  CURRENT FILING DATE:  2003-11-12
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 09/662,254
;  PRIOR FILING DATE: 2000-09-14
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
;   LENGTH: 32392
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-10-706-635-27

  Query Match             8.2%;  Score 41.2;  DB 22;  Length 32392;
  Best Local Similarity   48.7%;  Pred. No. 13;
  Matches  112;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy         14 TTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATT 73
              |||   |||||| |    |   |  ||  |  |||   |     |     || |||||| |   |||
Db       4783 TTATACATATAACTTAATATAATTATATTAATATTATTATATAATAAATATGGAAAATTATC 4842

Qy         74 TGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAG 133
                | |    | |||         | |  |     |    |  | ||||||  | |    |  |    |    |||
Db       4843 ATATTATTATATTAACAATTAAAAGAAATTCTGACAGATTACAAAAACTAGAAAATATAT 4902

Qy        134 ACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAA 193
                |        ||| |  |          |||||    || | || |  |   || |  || | |||||  |
Db       4903 TATCTTGTCAAAATTTATTATATAATAAAGATTATAGTGTATTTTATGGAATAGATTATA 4962

Qy        194 GTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAA 243
                 |   |  ||| ||       ||     || ||||      || ||   |  |    ||
Db       4963 AAAATATAAATAAAAATAATTTAAAAAATATATGTAAAAAAGGATTTAAA 5012


RESULT 9
US-10-260-238-1655/c
; Sequence 1655, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
;  APPLICANT: Budworth, Paul R.
;  APPLICANT:  Moughamer, Todd G.
;  APPLICANT:  Briggs, Steven P.
;  APPLICANT:  Cooper, Bret
;  APPLICANT:  Glazebrook, Jane
;  APPLICANT:  Goff, Stephen A.
;  APPLICANT:  Katagiri, Fumiyaki
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Provart, Nicholas
;  APPLICANT:  Ricke, Darrell
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
;  FILE REFERENCE: 60111-NP
;  CURRENT APPLICATION NUMBER: US/10/260,238
;  CURRENT FILING DATE:  2002-09-26
;  PRIOR APPLICATION NUMBER: US 60/325,448
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: US 60/325,277
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: US 60/370,620
;  PRIOR FILING DATE: 2002-04-04
;  NUMBER OF SEQ ID NOS: 6077
```

```
; SEQ ID NO 1655
;   LENGTH: 2000
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-10-260-238-1655

  Query Match             8.1%;  Score 40.8;  DB 18;  Length 2000;
  Best Local Similarity   48.0%;  Pred. No. 5;
  Matches  117;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy          4 AGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATG 63
              | |  |||   ||   ||| || ||      ||  |||  |||      | || ||   |
Db       1937 ATTTCGTATGATAAACATAGAAAATAAATATAAGATGAAGATGTGTTCTTTCAACTAAAT 1878

Qy         64 GATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAA 123
              || ||   |||| || |   || |   | || ||  | | || |  ||  | |   | |
Db       1877 GACACGATTTTGTATGCTTCATATAATTGCATATTTTATAATCATCGAGAAAACTATAAG 1818

Qy        124 CCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCA 183
              | |  || |      |   | |  ||    ||  || || |    ||  ||| |
Db       1817 CAATTGAGATGATTTGCTTTATGATATTTTAGATCAAATATTGCTTGATCGTCTAAAGGA 1758

Qy        184 AGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAA 243
              | | | | |  | |       | | | || |||| |  ||| | | ||||| |||  ||
Db       1757 AAAAACTAAGATGATGGGACACAAAGGACAATGTGGACCATAGAGCTATAAATAATTAAA 1698

Qy        244 GTGG 247
               |||
Db       1697 ATGG 1694


RESULT 10
US-10-741-601-5619
; Sequence 5619, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STENOSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001500
;  CURRENT APPLICATION NUMBER: US/10/741,601
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 26415
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5619
;   LENGTH: 59914
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(59914)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-601-5619

  Query Match             8.1%;  Score 40.8;  DB 20;  Length 59914;
  Best Local Similarity   50.5%;  Pred. No. 22;
  Matches   99;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy         76 AATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGAC 135
              | |||||| | | | |  | |  || ||  || |       |   |  |  |   |||
Db      35392 ACTTCCAAGTATGTTTATAAAATATATGTCTAAAACCATTAAGGTTGAGAAAATATACTG 35451

Qy        136 ACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGT 195
                |||| |||  |||    |||| ||  | |||  | | ||||  || || ||||| |
Db      35452 CATGAAAAAAATTTTTAAGTAAATTAGAATACTGCAAATGCATAATAAACAGCTTTAAAT 35511

Qy        196 TTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAA 255
                | |      | |||    | ||  ||  || |||   ||  | |  || ||||
Db      35512 RGCTACCTGAGATGTCACTTGATGTTGATCAAAATATGAAAGGAGCTGGAGTTTAGTGAA 35571

Qy        256 CAGTCTCTTCTCTGAT 271
              ||   ||| |  || |
Db      35572 CACAGTCTCCAGTGTT 35587


RESULT 11
US-10-741-601-5661/c
; Sequence 5661, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STENOSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001500
;  CURRENT APPLICATION NUMBER: US/10/741,601
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 26415
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5661
;   LENGTH: 101782
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(101782)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-601-5661

  Query Match             8.1%;  Score 40.8;  DB 20;  Length 101782;
  Best Local Similarity   50.5%;  Pred. No. 27;
  Matches   99;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy         76 AATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGAC 135
              | |||||| | | | |  | |  || ||  || |       |   |  |  |   |||
Db      11322 ACTTCCAAGTATGTTTATAAAATATATGTCTAAAACCATTAAGGTTGAGAAAATATACTG 11263

Qy        136 ACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGT 195
                |||| |||  |||    |||| ||  | |||  | | ||||  || || ||||| |
Db      11262 CATGAAAAAAATTTTTAAGTAAATTAGAATACTGCAAATGCATAATAAACAGCTTTAAAT 11203

Qy        196 TTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAA 255
                | |      | |||    | ||  ||  || |||   ||  | |  || ||||
Db      11202 RGCTACCTGAGATGTCACTTGATGTTGATCAAAATATGAAAGGAGCTGGAGTTTAGTGAA 11143

Qy        256 CAGTCTCTTCTCTGAT 271
              ||   ||| |  || |
Db      11142 CACAGTCTCCAGTGTT 11127


RESULT 12
US-10-437-963-36653/c
; Sequence 36653, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 36653
;   LENGTH: 1886
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_40458C.1
US-10-437-963-36653
```

```
  Query Match             8.0%;  Score 40.4;  DB 20;  Length 1886;
  Best Local Similarity   53.9%;  Pred. No. 6.2;
  Matches   83;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy        114 AGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTAT 173
              | | |||  |||| | |||| | ||  || | || | |   || |   | ||| |||
Db        217 AAAAATTGCACCAATGCTAGAATCAGATGGAGAGCTGTTCAATAAAAGATATGCTGATAT 158

Qy        174 TTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATA 233
              |    | |||| | || || |||||  |  | | |  |||||      | | |||  | |
Db        157 ATACACATCAAGGGTTTCAAATTTTCTACATTATACTCCATTCATGTATTTCCAATCACA 98

Qy        234 AAAAATGCAAGTGGATATTGAACAGTCTCTTCTC 267
              | ||  | | |  |    ||  |||| |  | ||
Db         97 AGAACAGAATGCAGCATATGTTCAGTTTGCTATC 64


RESULT 13
US-10-424-599-18984
; Sequence 18984, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 18984
;   LENGTH: 797
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_117146C.1
US-10-424-599-18984

  Query Match             8.0%;  Score 40.2;  DB 19;  Length 797;
  Best Local Similarity   48.5%;  Pred. No. 4.9;
  Matches  111;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy         53 TTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGAT 112
              || ||  | |   || ||| || | ||  ||  |||| | ||  |||| |  ||| | ||
Db        293 TTGAATTTTTTTTTAAATAATTTATTTAGAATCTTTTATACATTTAATTTATTTATAAAT 352

Qy        113 TAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTA 172
              | || | | || ||    |         |  |||   | |    | |||    ||    ||
Db        353 TTGATAATTAAACAAATTTGATATTTAATTGAATGTTGTATTTAGATATTTATTACAATA 412

Qy        173 TTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAAT 232
              || |||   || | |  |  || ||  ||| || |    |||  ||  | ||  ||| |
Db        413 ATTAATAGTTAAAAAAACTTGGTATTTAATTGAATAGTGTATTTAGATGTTTATTACAGT 472

Qy        233 AAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAAT 281
              |   |       | |||   | |   ||   ||||   | || ||
Db        473 GATTGAAAATTAAAAAAATTTGGTATTTAATTGAATGATGTATTTAGAT 521


RESULT 14
US-10-473-126-240/c
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 240
;   LENGTH: 8056
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-240

  Query Match             8.0%;  Score 40;  DB 21;  Length 8056;
  Best Local Similarity   48.7%;  Pred. No. 15;
  Matches  109;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy         15 TAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTT 74
              || |  ||  |  | |   ||  ||  |  ||  ||  ||||||  |       |  |
Db       7159 TACGTTTACGAATTATATATAATATATTTTTATTAATATTTAAAAAAAATTATAAATTAA 7100

Qy         75 GAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGA 134
              || |  |||  ||  |    | ||  | |||| |  |||   | |  |  |  |
Db       7099 TAAATTAAAAAATTAATATTTAAAAAATAATTAAATTTTAATTAATTTAAAATTTTTTT 7040

Qy        135 CACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAG 194
                ||| |||| |  |   | |  |  ||| |  | |||  |||  || |    ||
Db       7039 TTTTTAAAAAAAAATACGAAAATAATATTTTTTTTTTTTTTTTTAATTAAAATCGATATA 6980

Qy        195 TTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAA 238
              ||||  || ||  |     ||| | ||| ||| || || ||||||
Db       6979 TTTTTTATATATAATCGTATTTTTAATAATAAAAAAAAAAAAA 6936


RESULT 15
US-10-674-124A-1055/c
; Sequence 1055, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
;  APPLICANT: INOKO, Hidetoshi
;  APPLICANT:  TAMIYA, Gen
;  TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
;  TITLE OF INVENTION:  GENETIC POLYMORPHISM MARKERS
;  FILE REFERENCE: ORIN-003CIP
;  CURRENT APPLICATION NUMBER: US/10/674,124A
;  CURRENT FILING DATE:  2003-09-26
;  PRIOR APPLICATION NUMBER: 10/257,511
;  PRIOR FILING DATE: 2003-03-07
;  PRIOR APPLICATION NUMBER: PCT/JP00/07621
;  PRIOR FILING DATE: 2000-10-30
;  PRIOR APPLICATION NUMBER: JP2000-112699
;  PRIOR FILING DATE: 2000-04-13
;  PRIOR APPLICATION NUMBER: JP2002-327516
;  PRIOR FILING DATE: 2002-09-28
;  PRIOR APPLICATION NUMBER: JP2002-383869
;  PRIOR FILING DATE: 2002-12-09
;  NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 1055
;   LENGTH: 350
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: chr1.fa.O7frz.118358138
;   FEATURE:
;   OTHER INFORMATION: Located on chromosome 1
;   FEATURE:
;   OTHER INFORMATION: Distance between a terminus base of telomere on
;   OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
;   OTHER INFORMATION: sequence : 111170513
;   FEATURE:
;   OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
;   OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
;   OTHER INFORMATION: 5'-terminus of this base sequence : 158826
US-10-674-124A-1055

  Query Match             7.9%;  Score 39.8;  DB 20;  Length 350;
  Best Local Similarity   50.8%;  Pred. No. 4.3;
  Matches   95;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;
```

```
Qy         51 CGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAG 110
              | | |||  |||  | ||||||||    |     | | | | | || ||| | ||  ||
Db        208 CATGTAACATATATACACATATTTTTTTATATATTATATATACATATATTTTGTATATTT 149

Qy        111 ATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGG 170
              |||| | ||  |   |   ||| | |   | | | |   | ||| |      |||  |
Db        148 ATTATATATATATATATATATATATATATATATATATTTCAACACAATTTCTTATTCCAG 89

Qy        171 TATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACA 230
              |  |||||    |      ||    | ||    ||     ||  |||||||| || |  ||
Db         88 TCATTCATCTATAGATACTTAGGTTGTTTTTATTTCTTGGCTGTTGTGAATAATGCTGCA 29

Qy        231 ATAAAAA 237
              || || |
Db         28 ATGAATA 22
```

Search completed: November 8, 2005, 23:53:31
Job time : 652.857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 22:16:08 ; Search time 2598.77 Seconds
(without alignments)
7367.447 Million cell updates/sec

Title: US-09-463-542-3_COPY_818_1320
Perfect score: 503
Sequence: 1 tagagtaagtaccttaggaa.....atatttggaaactgatgtct 503

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	51.6	10.3	1101	9	CNS00E47	AL068720 Drosophil
c	2	48.2	9.6	1101	9	CNS0039G	AL063921 Drosophil
	3	46.4	9.2	747	8	BH957102	odj02b02.
	4	45.6	9.1	1106	9	CNS04LKI	AL296235 Tetraodon
c	5	44.6	8.9	766	8	BH460886	BOGSK22TF
	6	43.4	8.6	843	9	CNS00CS1	AL059666 Drosophil
c	7	43.2	8.6	495	9	CC764640	CH240_47L
	8	43.2	8.6	618	9	CL335225	CL335225 RPCI44_25
c	9	43.2	8.6	1201	9	CNS0167M	AL106396 Drosophil
	10	42.6	8.5	623	8	BH742342	BH742342 gu74b05.g
	11	42.4	8.4	669	8	BZ465736	BZ465736 BONMQ35TF
c	12	42.4	8.4	818	8	BH661305	BOHYL59TF
c	13	42.2	8.4	766	8	AQ544187	AQ544187 RPCI-11-3
	14	42.2	8.4	874	9	CG019400	CG019400 ZUAAC48TV
c	15	42.2	8.4	1101	9	CNS000B8	AL063632 Drosophil
	16	42	8.3	473	6	CB888081	CB888081 taa87b01.
	17	42	8.3	959	9	CNS017OR	AL108309 Drosophil
	18	42	8.3	1101	9	CNS008X3	AL052544 Drosophil
c	19	41.8	8.3	266	9	CL883428	CL883428 abf62e12.
	20	41.8	8.3	496	8	BH711021	BH711021 BOMFC41TF
	21	41.8	8.3	591	8	BH467264	BH467264 BOGMM54TR
c	22	41.4	8.2	448	7	CK732164	CK732164 TgESTzyj7
c	23	41.4	8.2	617	9	CL625152	CL625152 OR_BBa002
c	24	41.4	8.2	628	9	CL611954	CL611954 OR_BBa000
c	25	41.2	8.2	1174	9	CL110860	CL110860 ISB1-54F2
	26	41	8.2	724	7	CV464338	CV464338 taj19b03.
	27	41	8.2	841	9	CR812616	CR812616 GR0AAA36D
c	28	41	8.2	1086	9	CNS00YXK	AL096962 Drosophil
c	29	41	8.2	1123	8	CC273164	CC273164 CH261-81O
c	30	40.8	8.1	321	4	BG438771	BG438771 dab28d09.
c	31	40.8	8.1	412	5	BP623936	BP623936 BP623936
c	32	40.8	8.1	538	5	BU811789	BU811789 UL88TE11
	33	40.8	8.1	693	9	CL934409	CL934409 OA_ABa004
	34	40.8	8.1	777	6	CB180244	CB180244 AGENCOURT
	35	40.8	8.1	832	8	BZ601497	BZ601497 WHACS32TF
c	36	40.8	8.1	1101	9	CNS00EJ4	AL069257 Drosophil
	37	40.8	8.1	1101	9	CNS00ZB7	AL097453 Drosophil
	38	40.6	8.1	705	8	BH970382	BH970382 odh08h10.
	39	40.6	8.1	734	9	CNS010MP	AL099163 Drosophil
	40	40.6	8.1	736	8	BZ469037	BZ469037 BONGG60TF
	41	40.6	8.1	1101	9	CNS00Z6Z	AL097301 Drosophil
	42	40.4	8.0	515	2	BE238266	BE238266 894041C12
	43	40.4	8.0	738	6	CA104351	CA104351 SCJFHR103
c	44	40.4	8.0	784	4	BI738064	BI738064 603357159
c	45	40.2	8.0	326	4	BG370931	BG370931 dac46d02.

ALIGNMENTS

RESULT 1
CNS00E47
LOCUS CNS00E47 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR28M18 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL068720
VERSION AL068720.1 GI:4948863
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR28M18"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

Query Match 10.3%; Score 51.6; DB 9; Length 1101;
Best Local Similarity 29.8%; Pred. No. 0.011;

```
  Matches   81;  Conservative   73;  Mismatches  118;  Indels    0;  Gaps    0;

Qy         70 TATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGG 129
              |:| | |:|| :|:|||| || :|::|:: |||::| :|| : :: |: :
Db        750 TRTGTAAMTTTTWAMTTTTGATTTWGMWARKARTATTWKATRTTTTTRGGRMMCTAARWK 809

Qy        130 ATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGAT 189
              |:| :    :| | ||:| : |     ||:| ||| | :   |: | :: :||
Db        810 ADACMTTTTTWATGNAAMCMYMGAGTGKATWAMWTTTGGAGAVMAAAABTTGTHRYTKAT 869

Qy        190 TTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGAT 249
              |||: |:|: :|:|:    :  :   ||  : |:| : | | | ::: :: :  |
Db        870 TTAHTTWTKATWTWTWYCCGRNTCWSTAGAGACMRASACMSASASACMMYMTHMMAKAYT 929

Qy        250 ATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCCCTCACGAGAC 309
              :|:  :|:|:  : ::| : |   :  :: ||::: ::::   :|  | :   | ||||
Db        930 WTWAKWCRGHTAYAKYTTKTAMTSKMTWWYATWMHMYHMMCMCMAAACASAGAARGAGAR 989

Qy        310 ACTGAACATGTGGTCACCGGCGAGACAGTGTG 341
              :  ::|   |:||     :| ||:| :| | |
Db        990 RRGRRARGGGKGGAAGARRGRGARAGRGGGRG 1021


RESULT 2
CNS0039G/c
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:7227"
                      /clone="BACR08K10"
                      /clone_lib="RPCI-98"
                      /note="end : TET3"
ORIGIN

  Query Match             9.6%;  Score 48.2;  DB 9;  Length 1101;
  Best Local Similarity   15.7%;  Pred. No. 0.084;
  Matches   44;  Conservative  139;  Mismatches   98;  Indels    0;  Gaps    0;

Qy          7 AAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGAT 66
              :|: :  |::  |::::|: ::: | ::: |:    :||  :: |:    :  | ::::
Db        860 WADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRD 801

Qy         67 ACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCA 126
              : : ::: :||::  :|||:||: ||  ::|:::::: :|: : | :: ::::: ::  :
Db        800 RKRADDKRDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDR 741

Qy        127 GGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGA 186
              | | ::: :  :::::::::  :::    : | :::     :| :| :    | :: ::::
Db        740 AGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRD 681

Qy        187 GATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTG 246
              :||::: : ::   |::|  : :   :::::::::: | : ::  | ||::::|  |:::
Db        680 RATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWA 621

Qy        247 GATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTA 287
              :::::|:|: ::    :| | :  :::: :|:|:: ::::
Db        620 KWDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDR 580


RESULT 3
BH957102
LOCUS       BH957102                747 bp    DNA     linear   GSS 01-OCT-2002
DEFINITION  odj02b02.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH957102
VERSION     BH957102.1  GI:23438329
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
            Nash,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: odj02  row: b  column: 02
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 8
            High quality sequence stop: 551.
FEATURES              Location/Qualifiers
     source           1. .747
                      /organism="Brassica oleracea"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:3712"
                      /clone_lib="B.oleracea002"
                      /note="Vector: pOTw13; Whole genome shotgun library from
                      flowering buds. DNA was purified  from a crude nuclear
                      prep using Brassica oleracea TO1000DH3 buds  provided by
                      Thosmas Osborn at the University of Wisconsin. Genomic
                      DNA was provided by Pablo Rabinowicz (CSHL) and the
                      shotgun library  prepared at Washington University Genome
                      Sequencing Center."
ORIGIN

  Query Match             9.2%;  Score 46.4;  DB 8;  Length 747;
  Best Local Similarity   52.6%;  Pred. No. 0.23;
  Matches  101;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy         10 TACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACA 69
              ||   ||  |||| || | |     |   | | |||  ||  |  |||      |||||
Db         51 TAATATATTAATAAAATAATATTTAACGTTTTTTATATTAATATAAAAATATGATATACA 110

Qy         70 TATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGG 129
              | |||||||| |||| || || | ||||     ||||    | ||| ||  |   |
Db        111 CAATTGAATTCAAAATAATTATTTAAATAAATACATTAATAAAAAGAAATGAATATATTA 170
```

```
Qy        130 ATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGAT 189
              ||| | |   ||  || |    |  || | ||    |    | ||| |||||    | ||
Db        171 ATAAAAATTTAACTAACATAAAGGGCACAAAAAAAATAATTTTTTTTATAAGTTGTACAT 230

Qy        190 TTAAGTTTTCCA 201
              |||| ||| | |
Db        231 TTAAATTTACAA 242
```

RESULT 4
CNS04LKI
LOCUS CNS04LKI 1106 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 118P23 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL296235
VERSION AL296235.1 GI:8034815
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3 (bases 1 to 1106)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .1106
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="118P23"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG118CH12LP1~end : T7"
ORIGIN

Query Match 9.1%; Score 45.6; DB 9; Length 1106;
Best Local Similarity 41.4%; Pred. No. 0.41;
Matches 116; Conservative 24; Mismatches 140; Indels 0; Gaps 0;

```
Qy         20 ATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTTGAATT 79
              | | || |:|:   |    | |   || |    ||| |  |: | :  | |:: ||
Db        777 AAAAAAAAWTWTTWTTTTTTTCCTTTAAAACAACTTACAACAWCAAAWTTTTTYWTAAAA 836

Qy         80 CCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCG 139
                ||||:  ||:|||| | ||| | |  | | |  |   :| : | ||   | |   |||
```

```
Db        837 ACAAAWAATTYTTCACACAATTTWACAAAAAAAAATTWWTAYCATCATAAAAAACAACCT 896

Qy        140 AAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTC 199
                |||||||:     |||: :  ||  |   |||  |||      |   || ::||
Db        897 YYTAAAAACTTYTTAAAAAHYWTTTTAWTWACCTTTTTTAACTTYYTAAAAAAACYYTTT 956

Qy        200 CATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAACAGT 259
                |    |  | :| |  :  |   ::||| |  ||||||:  || |  |: ||     |
Db        957 TAAAACACCACYCKTCCYACAAACWWCAAAATAAAAAAAWWTAATTAYAWTTTTYTTTTT 1016

Qy        260 CTCTTCTCTGATAATTCTAAATACAGTACAGTTCACGCCC 299
                | :| | |  | |:| ||  :| |  | |    || |||
Db       1017 TTTBTTTTTTTTTAWTTTATTWAAAWAAAACAAAACACCC 1056
```

RESULT 5
BH460886/c
LOCUS BH460886 766 bp DNA linear GSS 13-DEC-2001
DEFINITION BOGSK22TF BOGS Brassica oleracea genomic clone BOGSK22, genomic survey sequence.
ACCESSION BH460886
VERSION BH460886.1 GI:17650631
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 766)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOGSK22TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .766
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGSK22"
/clone_lib="BOGS"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Query Match 8.9%; Score 44.6; DB 8; Length 766;
Best Local Similarity 49.4%; Pred. No. 0.7;
Matches 116; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

```
Qy         27 ATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATT 86
              ||  |  ||   |  || ||  ||  | | || | |  || || | |  ||||    ||
Db        708 ATAACTATATATTTTTGTTATAAATTTATTAAGTGTTTATTCAAAATGTAATTAACTATA 649

Qy         87 TTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAA 146
              |   | | ||| | |  |||| || | | ||  ||| |    || | |   ||  |||
Db        648 ATATATAACATATTCTTGTTAGATATAACATATGTAACAACTTGTAAAAATTAAATAAAA 589

Qy        147 ACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAA 206
              |  |  |||||   || ||  |  | | |   |  |||| |  | | ||| |
Db        588 CAATGCCGAAAATATTTTTGCTTCAAAATTTTCATTTTGTTATTTTATATATACATATTT 529

Qy        207 GAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAACAGTCT 261
              | |  | | |  || |   |  |  |  |||   |  ||  |||||||  | |
Db        528 TAGGAAAATATACATAAACTATGAGGACCAAACTAATTTGATTATTGAACTTTTT 474
```

```
RESULT 6
CNS00CS1
LOCUS       CNS00CS1                 843 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR26H19 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL059666
VERSION     AL059666.1  GI:4947129
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR26H19"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             8.6%;  Score 43.4;  DB 9;  Length 843;
  Best Local Similarity   36.7%;  Pred. No. 1.5;
  Matches   90;  Conservative   39;  Mismatches  116;  Indels    0;  Gaps    0;

Qy         38 ATGCTGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGAT 97
              :|  | |||  |   |||::: | |  |:| | | :| ||||  | |:| || |    :|
Db        578 WTTTTTATATWATTTTTTWWWTTTTTTAYAWAAAAWTTAATTWTATAWTATTWTAATAWT 637

Qy         98 AATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAA 157
              : | : |::| : |:|| : |:  ||  |   ||| : :   :|| ||    :|   ::
Db        638 WTTAWTAWWAAWAAWTATWAAWATAAAAWATAAATAWWWWWTWWAATAATTAAWTTTATWW 697

Qy        158 ATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGT 217
                |  | |  |||:  : |   || | || :|  : :     : ||  |   ||  :
Db        698 TATATATATWTAATATWAWWAATTWAAWATATAWATAWAWATATAWAAATTAATAATAWW 757

Qy        218 GAATTATACAACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCT 277
               ||  |:|  | |||||||:||   |  :  :|||| :: | :    |      | |: : |
Db        758 AAAAAAWAWTAAAATAAAWAAATWATAYAWWTATTHWWAAWHAAAATTATAAAAAWAWAT 817

Qy        278 AAATA 282
              | ||:
Db        818 ATATW 822


RESULT 7
```

```
CC764640/c
LOCUS       CC764640                 495 bp    DNA     linear   GSS 27-JUN-2003
DEFINITION  CH240_47L12.TJ CHORI-240 Bos taurus genomic clone CH240_47L12,
            genomic survey sequence.
ACCESSION   CC764640
VERSION     CC764640.1  GI:32311143
KEYWORDS    GSS.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.,
            Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L.,
            Womack,J.E., de Jong,P.J. and Lewin,H.A.
  TITLE     Bovine BAC end sequences from CHORI-240 library
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CH240_47L12.TV
            Contact: Harris Lewin
            Department of Animal Sciences
            University of Illinois at Urbana Champaign
            1201 W. Gregory Dr., Urbana, IL 61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5617
            Email: h-lewin@uiuc.edu
            Clones are derived from the bovine BAC library CHORI-240
            (http://www.chori.org/bacpac/bovine240.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/ordering_information.htm). This work
            was undertaken as part of the International Bovine BAC Mapping
            Consortium (IBBMC) by by University of Illinois at Urbana
            Champaign, USA with funds provided by grant No. AG202-34480-11828
            from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock
            Genome Sequencing Initiative)
            Plate: 47  row: L  column: 12
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /strain="breed: Hereford"
                     /db_xref="taxon:9913"
                     /clone="CH240_47L12"
                     /sex="Male"
                     /cell_type="Blood"
                     /clone_lib="CHORI-240"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
                     library (Male) produced by Pieter de Jong"
ORIGIN

  Query Match             8.6%;  Score 43.2;  DB 9;  Length 495;
  Best Local Similarity   50.0%;  Pred. No. 1.5;
  Matches  108;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy         21 TATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTTGAATTC 80
              | ||| ||||||  |          | |   |  | |  |  ||    | | || | || |
Db        452 TTTAAAATTTCAAAAAATGCACAAAATACAAATATGTAAAATTTTGAAACATCTTAAATT 393

Qy         81 CAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGA 140
              ||||  |  | || | |||||||  ||  | |    || |||   ||| |||   || ||
Db        392 CAAAAGTAACATCTGTTAATGTGGGTAATGCTGCTGGAGTCATTTAGGTATATCAACAGA 333

Qy        141 AAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCC 200
              |||| | | |     ||||| |||||      |  || |       | || ||  | |  |
Db        332 GAGAAGAATATATATAAATATGCTTTTATATGATTGAGCTCTTATAGGTTCTATTCCTAT 273

Qy        201 ATTTAAGAAGCCATTGTGAATTATACAACAATAAAA 236
              || ||| |       | |||  |  |  |||  ||
Db        272 TTTCAAGTATTTCAATTTAATAGTGTATAAATTTAA 237
```

```
RESULT 8
CL335225
LOCUS       CL335225                618 bp    DNA     linear   GSS 19-AUG-2004
DEFINITION  RPCI44_253A15.r RPCI-44 Sus scrofa genomic clone RPCI44_253A15,
            genomic survey sequence.
ACCESSION   CL335225
VERSION     CL335225.1  GI:51387193
KEYWORDS    GSS.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   Rogatcheva,M.B., Meyers,S., He,W., Larkin,D.M., Marron,B.M.,
            Beever,J.E. and Schook,L.B.
  TITLE     Piggy-BACing the Human Genome: Constructing a Porcine Physical Map
            Through Comparative Genomics
  JOURNAL   Unpublished (2004)
COMMENT     Other_GSSs: RPCI44_253A15.f
            Contact: Lawrence B. Schook
            Department of Animal Sciences
            University of Illinois at Urbana Champaign
            1201 W. Gregory Dr., Urbana, IL 61801, USA
            Tel: 217 265 5326
            Fax: 217 244 5617
            Email: schook@uiuc.edu
            Clones are derived from the porcine BAC library RPCI-44
            (http://www.bacpac.chori.org/porcine242.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@chori.org).
            Clones may be purchased from BACPAC Resources
            (http://BACPACorders.chori.org). This work was undertaken as part
            of the International Swine Genome Sequencing Consortium by
            University of Illinois at Urbana Champaign, USA with funds provided
            by grant No. AG2002-34480-11828 from USDA-CSREES and
            AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing
            Initiative)
            Plate: 253  row: A  column: 15
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .618
                     /organism="Sus scrofa"
                     /mol_type="genomic DNA"
                     /strain="four pigs (breed: 37.5% Yorks Landrace and 25%
                     Meishan)"
                     /db_xref="taxon:9823"
                     /clone="RPCI44_253A15"
                     /sex="male"
                     /cell_type="blood"
                     /clone_lib="RPCI-44"
                     /note="Vector: pTARBAC2; Site_1: EcoRI; Site_2: EcoRI;
                     porcine male BAC library produced by Pieter de Jong"
ORIGIN

  Query Match             8.6%;  Score 43.2;  DB 9;  Length 618;
  Best Local Similarity   48.3%;  Pred. No. 1.6;
  Matches  117;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy         53 TTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGAT 112
              ||||   ||  |||  | |   ||| ||||  |     ||  ||| | |||  |
Db        357 TTTATTTCATTTTTACCCAATCTCATTTCAAAAATAATCAGAGGGAATTTTATTTAAATA 416

Qy        113 TAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTA 172
              |  | || | | | | |||| |    ||     |       ||| |   | ||  | ||
Db        417 TGTAATTTTAGCTAAGTATAGTTAATATAAAGTCAACACAAGCAACTGGCCATTGAGTTA 476

Qy        173 TTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAAT 232
               ||   ||   ||| |   |  ||    ||| |  |    || | |  ||||||  ||
Db        477 CTTATAAATTCAGAAAAAAAGCTTACTTATTGCATTATTACTTTTCATCTATACAGCAGA 536

Qy        233 AAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTT 292
              ||| |||| |  |   |||   | | |  |  || ||| || |   | | ||  |  |||
Db        537 AAACAATGTACATATTTATCTCAAATTTGCAACTATGAAAAGTGCTATTCCATAANNGTT 596

Qy        293 CA 294
              ||
Db        597 CA 598


RESULT 9
CNS0167M/c
LOCUS       CNS0167M                1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106396
VERSION     AL106396.1  GI:5621701
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15M24"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             8.6%;  Score 43.2;  DB 9;  Length 1201;
  Best Local Similarity   41.5%;  Pred. No. 1.8;
  Matches  102;  Conservative   22;  Mismatches  122;  Indels    0;  Gaps    0;

Qy         44 ATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTG 103
              | |  |:  | |:|| : |   |  ||:| | ||:   :  ||||: || | | ||
Db        949 AAAAAAWAATATWAAAWTTTTTTTTATWTATAAAWAWAWTTTTTTYWTTAAAAAAAAAW 890

Qy        104 ATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGC 163
              | | |  :|| | |||  |  |   | : | |   ||| ||||  ::    |: | :
Db        889 TTAAAAATWTAAAAATTATAAAATAAAAWAAAAAAAAAAAAAAAAAAWWAAAAAWTTTWAT 830

Qy        164 TTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTA 223
              | |   | ||| | ||  || | |||||| |:||      || ||   |:   |:::|
Db        829 TATAAWATTTTTWAAAAAAAAAAAATTTAATTYTTTTTNAAAATAAAAAAWAAAWAWWWA 770

Qy        224 TACAACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATAC 283
              | || ::|||| |           ||  ||||     :  | : ||||    ||| |
Db        769 AAAAATWWTAAATATAAWTTAAATNCATAAAACAAAAAAWAATTWHATAAAAAAAAAAAA 710

Qy        284 AGTACA 289
              | | |
Db        709 AAAAAA 704


RESULT 10
```

```
BH742342
LOCUS       BH742342                623 bp    DNA     linear   GSS 25-FEB-2002
DEFINITION  gu74b05.g1 BoBuds01 Brassica oleracea genomic clone gu74b05 5',
            genomic survey sequence.
ACCESSION   BH742342
VERSION     BH742342.1  GI:18876955
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Katari,M., O'Shaughnessy,A., Palmer,L., Bahret,A., Baker,J.,
            Balija,V., Cunnius,D.M., Katzenberger,F., King,L., Kirchoff,K.,
            Kuit,K., Miller,B., Muller,S., Nascimento,L., Preston,R.,
            Santos,L., Shah,R., Zutavern,T., Dedhia,N., Rabinowicz,P.D. and
            McCombie,W.R.
  TITLE     Whole Genome Shotgun Reads from Brassica oleracia (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: gu74  row: b  column: 05
            Seq primer: -21UnivRev
            Class: shotgun
            High quality sequence stop: 623.
FEATURES             Location/Qualifiers
     source          1. .623
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone="gu74b05"
                     /clone_lib="BoBuds01"
                     /note="Vector: M13 for .x reads, pBluescript for .b and .g
                     reads; Site_1: EcoRV; Whole genome shotgun library from
                     flowering buds. DNA was purified from a crude nuclear prep
                     using Brassica oleracea TO1000DH3 buds provided by Thomas
                     Osborn at the University of Wisconsin. Genomic DNA
                     provided by Pablo Rabinowicz (CSHL) and shotgun library
                     prepared in McCombie Lab."
ORIGIN

  Query Match             8.5%;  Score 42.6;  DB 8;  Length 623;
  Best Local Similarity   47.5%;  Pred. No. 2.3;
  Matches  126;  Conservative    0;  Mismatches  139;  Indels    0;  Gaps    0;

Qy          11 ACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACAT 70
               ||||||  ||||  |  || ||  |  ||  | |     |  |  |||  | | | ||| |
Db         192 ACCTTATAAATAGCAATTTCCATAAAAATTATTACGAAATTGCCTAATATTTAGCTACCT 251

Qy          71 ATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGA 130
               | |  | ||  ||  | | |      | || | |||  |    | |    ||   |
Db         252 AATAAATTTAGAAGATATGCAATCCTTCATTTAATTTAAATCAAAATCAAAAATTCGATT 311

Qy         131 TAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATT 190
               | |  | | | | |||||| |      |       || |   |||   | |       |||
Db         312 TTGCAATCCATAAAAAACTATAAAATACATTAAATTATCAGATTAATTCATTGGTTTATT 371

Qy         191 TAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATA 250
                ||  ||     |      ||  ||||   | || ||||| ||||||||   |      |||
Db         372 GAAAATTAAATTAATCAAAATCATTTAAATTTTTACAATAATAAAAATAACTCAGTTATA 431

Qy         251 TTGAACAGTCTCTTCTCTGATAATT 275
               || ||     |  |||||  |||||
Db         432 TTTAATTAATTAGTCTCTTCTAATT 456


RESULT 11
```

```
BZ465736
LOCUS       BZ465736                669 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONMQ35TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONMQ35,
            genomic survey sequence.
ACCESSION   BZ465736
VERSION     BZ465736.1  GI:26757807
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BONMQ35TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONMQ35"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             8.4%;  Score 42.4;  DB 8;  Length 669;
  Best Local Similarity   50.5%;  Pred. No. 2.6;
  Matches  103;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy           6 TAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGA 65
               ||  |||||||  |          || |  |  ||  | | || ||  |||| ||| |  |
Db         451 TAGTTACCTTAAAATATATTTTGTTTAAAATAATCATAAAACAAAATTTTACACTTTATA 510

Qy          66 TACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACC 125
               |||||||||  || || ||||  |  || | |      |   ||    |  |  |||||||
Db         511 TACATATTTCAAATCAAAATAATAATTTAAAATTGATTTATATCAAAAATTGATTCAAAA 570

Qy         126 AGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAG 185
               |  | | |    ||| || |  ||   | || |   ||||    ||      ||   ||
Db         571 ATATACATATGTTTAAAAAATAAATTTTACTAAAATATTTTCCAATAACCATTATTAAAA 630

Qy         186 AGATTTAAGTTTTCCATTTAAGAA 209
               | ||         |  || ||||||
Db         631 AAATGGTTTCAATATATATAAGAA 654


RESULT 12
BH661305/c
LOCUS       BH661305                818 bp    DNA     linear   GSS 19-FEB-2002
DEFINITION  BOHYL59TF BO_2_3_KB Brassica oleracea genomic clone BOHYL59,
            genomic survey sequence.
ACCESSION   BH661305
VERSION     BH661305.1  GI:18720067
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 818)
```

```
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .818
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHYL59"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             8.4%;  Score 42.4;  DB 8;  Length 818;
  Best Local Similarity   47.7%;  Pred. No. 2.7;
  Matches  124;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy         28 TTTCAGTAGCATGCTGATACCAACGTTTAAACTATGGATACATATTTGAATTCCAAATTT 87
              ||||   |  |  |  | | |||  |||||||| |  |||||||||| || || ||||
Db        301 TTTCTAAACTAACCATAAAACAAAATTTAAACTTTATATACATATTTCAAATCAAAATAA 242

Qy         88 TTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAGAAAA 147
              |  |||| |      |   ||   |  | |||||||  |   ||| | |     | |||
Db        241 TAATTCAAAGTTGATTTATATCAAAAATTGATTCAAAAATATATATATATATTCAAAAAT 182

Qy        148 CTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATTTAAG 207
                | | |   ||  || |  ||  |    || | || | |   |  || ||||
Db        181 GGATTTCTACTAAACTATTTTTCAATAACCATTAAAAAAAATGTTATCAATATATATAAG 122

Qy        208 AAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAACAGTCTCTTCTC 267
              ||     |   |   | | |  | ||| |   |     || ||     | | || |
Db        121 AAAAATATAATACAGAGCCCAATTTGAAATACCAACTCAAATTATGGTTTTTATATTTTA 62

Qy        268 TGATAATTCTAAATACAGTA 287
              |  ||| | |  | |   ||
Db         61 TATTAAGTTTTTAAAATATA 42


RESULT 13
AQ544187/c
LOCUS       AQ544187                 766 bp    DNA     linear   GSS 19-MAY-1999
DEFINITION  RPCI-11-326P5.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-326P5,
            genomic survey sequence.
ACCESSION   AQ544187
VERSION     AQ544187.1  GI:4868489
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 766)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-326P5.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .766
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7625164"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-326P5"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             8.4%;  Score 42.2;  DB 8;  Length 766;
  Best Local Similarity   49.8%;  Pred. No. 3;
  Matches  107;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy         59 CTATGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGA 118
              |||| | |  |||    ||||    |||||| ||   ||||| | | |||| ||     |
Db        526 CTATTGCTGTATAAACAAATTATTTATTTTTGTTACTATAATATCACTAGAAATATATTA 467

Qy        119 TTCAACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCAT 178
              || | ||  ||  || |  |      |||||   |  |   |||  |    |   || ||
Db        466 TTTATACAAACATTTATCCATTGTGAAAATGGTAATCTTTTAATAGTCTCTGAGATTAAT 407

Qy        179 AAGCAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAA 238
              ||  |  ||||  | | | |  || |   ||    |  ||| ||||| |   |||||
Db        406 AATTAGCAGATCAAGGCTATAAATATTTTTAGGGCTCTTGATTTATATATTGATAAATTG 347

Qy        239 TGCAAGTGGATATTGAACAGTCTCTTCTCTGATAA 273
                | |       ||  |||| | |  | ||  || |
Db        346 CTCCACAAATAATATAACAATTTAGTGTCACATTA 312


RESULT 14
CG019400
LOCUS       CG019400                 874 bp    DNA     linear   GSS 19-AUG-2003
DEFINITION  ZUAAC48TV ZM_3.0_4.0_KB Zea mays genomic clone ZMMBPa0005G23,
            genomic survey sequence.
ACCESSION   CG019400
VERSION     CG019400.1  GI:33891565
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 874)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: ZUAAC48TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
```

```
FEATURES             Location/Qualifiers
     source          1. .874
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBPa0005G23"
                     /clone_lib="ZM_3.0_4.0_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 3-4 kb 'unfiltered'
                     genomic DNA library"
ORIGIN

  Query Match             8.4%;  Score 42.2;  DB 9;  Length 874;
  Best Local Similarity   50.8%;  Pred. No. 3.1;
  Matches  101;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy          84 ATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTCAACCAGGGATAGACACCGAAAG 143
               ||| |  |    |  ||| | || | | |||| |     |  |     |    ||||
Db         389 ATTAATTATTTTCTGTTGTAAATAAATAATAGAAAAAATGCATGCTTAAATATTGGAAAA 448

Qy         144 AAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAGCAAGAGATTTAAGTTTTCCATT 203
               || | |||  |     | |  ||  |||  | ||| |  |    |  |||  |||||   |
Db         449 AATATTTTTACACCCGATGTGTTACGGTTATGCATTAATATTGGTATTATTTTTTCTGCT 508

Qy         204 TAAGAAGCCATTGTGAATTATACAACAATAAAAAATGCAAGTGGATATTGAACAGTCTCT 263
               | |     |    || |||  || || | ||  |||||| |||||   |  |  || | |
Db         509 TTATTTTCTGAAGTAAATAGTAGAAAATTATTAAATGCATGTGGAGTATATATTGTGTAT 568

Qy         264 TCTCTGATAATTCTAAATA 282
                 ||   ||   ||||||||
Db         569 ACTTAAATCGCTCTAAATA 587


RESULT 15
CNS000B8/c
LOCUS       CNS000B8               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR01A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063632
VERSION     AL063632.1  GI:4938680
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR01A24"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             8.4%;  Score 42.2;  DB 9;  Length 1101;
  Best Local Similarity   37.1%;  Pred. No. 3.3;
  Matches  108;  Conservative   37;  Mismatches  146;  Indels    0;  Gaps    0;

Qy           2 AGAGTAAGTACCTTAGGAATATAACATTTCAGTAGCATGCTGATACCAACGTTTAAACTA 61
               | |  ||  |  ||:  |    || ||:: | :|  ||  :  ||  ||  || ||: ||
Db        1029 AAAAAAATAAAATTWTAATATAAAAATWWTAAWATAATATWTTTAATAAAWTTAAAWATA 970

Qy          62 TGGATACATATTTGAATTCCAAATTTTTCTTCAGATAATGTGATTAGAGATTAGAGATTC 121
               :  :|| | |     ||     ||          | ||||   | :| | :  : | :   |
Db         969 WATWTAWAAAAAAAAAAAAAATAAAAAAAAAAWAAATAAAWTTWTAAWWATWAAAWATATA 910

Qy         122 AACCAGGGATAGACACCGAAAGAAAACTTTGCCCAAATAAGCTTTCTGGTATTTCATAAG 181
               |    |   ||: | |   ||| :     :|    || | |  ||| |  ||:|  | ||
Db         909 ATTTAAWWATWAAAAATWAAAAWTTTAAWTWWTAAATTTAATTTTTTTTTAWTAAAAAAA 850

Qy         182 CAAGAGATTTAAGTTTTCCATTTAAGAAGCCATTGTGAATTATACAACAATAAAAAATGC 241
               || : |   || |      |   |: ||   : | : :| :||| :| | :|||||:
Db         849 AAATWAAAAAAAATAAATWAAAAAWWAAAWWWATAWAWAAWATAAWAAATWAAAAAAHTA 790

Qy         242 AAGTGGATATTGAACAGTCTCTTCTCTGATAATTCTAAATACAGTACAGTT 292
               || :  |  | :| | |::  | :   |||| :  ||:|: : :| | ::
Db         789 AATWAAAATTACMAAAATYWAATTYAAAATAAAWWAAAWTWTWAWAWAWWW 739


Search completed: November  9, 2005, 04:17:50
Job time : 2602.77 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:46:48 ; Search time 101.246 Seconds
(without alignments)
7308.644 Million cell updates/sec

Title: US-09-463-542-1_COPY_1_125
Perfect score: 125
Sequence: 1 cccctgcccctgcccctgcc.....gcgggcaggcggggcccagc 125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	125	100.0	503	2	AAX19032	Aax19032 Human PPA
c	2	118.6	94.9	12548	13	ADS36487	Ads36487 Human aut
	3	118.6	94.9	158417	13	ADS36461	Ads36461 Human aut
	4	81.4	65.1	201	2	AAX19063	Aax19063 Human PPA
c	5	57.8	46.2	1416	8	ABZ20967	Abz20967 Animal te
c	6	56.6	45.3	420	13	ACN51217	Acn51217 Cotton an
	7	54.4	43.5	1315	2	AAZ41312	Aaz41312 Human nor
	8	54.2	43.4	200	5	ABV61578	Abv61578 Human pro
c	9	54.2	43.4	629	13	ACN54594	Acn54594 Cotton an
	10	53	42.4	1131	12	ADQ23807	Adq23807 Human sof
c	11	52.2	41.8	168	6	ABN69641	Abn69641 Streptoco
c	12	52	41.6	3163	10	ADC87060	Adc87060 Human GPC
	13	52	41.6	28198	10	ADG37080	Adg37080 Mouse pla
c	14	51.8	41.4	344	4	AAI82007	Aai82007 Human pol
c	15	51.8	41.4	437	13	ACN58918	Acn58918 Cotton gy
c	16	51.4	41.1	110000	12	ADQ97050_0	Adq97050 Human can
c	17	51.2	41.0	349	4	AAI84504	Aai84504 Human pol
c	18	51	40.8	588	13	ACN54596	Acn54596 Cotton an
	19	51	40.8	600	6	ABQ52497	Abq52497 Oligonucl
c	20	51	40.8	600	6	ABQ52496	Abq52496 Oligonucl
	21	51	40.8	2188	2	AAZ77506	Aaz77506 Human ova
	22	50.8	40.6	142	7	ADR41362	Adr41362 Human CD-
c	23	50.8	40.6	1045	6	ABI99807	Abi99807 Mouse isc
c	24	50.8	40.6	1286	6	ABI99656	Abi99656 Mouse isc
c	25	50.6	40.5	434	4	AAI84343	Aai84343 Human pol
	26	50.6	40.5	3198	2	AAX02974	Aax02974 Human IL-
c	27	50.4	40.3	594	6	ABQ43958	Abq43958 Oligonucl
	28	50.4	40.3	594	6	ABQ43959	Abq43959 Oligonucl
	29	50.2	40.2	1845	3	AAZ98400	Aaz98400 Canine be
	30	50.2	40.2	1845	6	ABK40732	Abk40732 Dog beta1
c	31	50.2	40.2	43058	6	ABL64982	Abl64982 Lung canc
c	32	50.2	40.2	43058	6	ABL65219	Abl65219 Lung canc
c	33	50.2	40.2	43058	6	ABN97455	Abn97455 Gene #395
	34	50	40.0	615	8	ACA23975	Aca23975 Prokaryot
c	35	50	40.0	1948	6	ABQ69181	Abq69181 Listeria
c	36	50	40.0	5082	6	ABQ70953	Abq70953 Listeria
c	37	49.8	39.8	204	13	ADR93549	Adr93549 Novel S.
	38	49.8	39.8	209	13	ADR91550	Adr91550 Novel S.
	39	49.8	39.8	282	13	ADR91549	Adr91549 Novel S.
c	40	49.8	39.8	308	13	ADR93088	Adr93088 Novel S.
c	41	49.8	39.8	612	13	ADR93335	Adr93335 Novel S.
	42	49.8	39.8	884	8	ACC48196	Acc48196 EMCV inte
c	43	49.8	39.8	1387	10	ADC87462	Adc87462 Human GPC
c	44	49.8	39.8	2685	6	ABK83742	Abk83742 Human cDN
c	45	49.8	39.8	12733	6	ABK98631	Abk98631 Vector pE

ALIGNMENTS

RESULT 1
AAX19032
ID AAX19032 standard; DNA; 503 BP.
XX
AC AAX19032;
XX
DT 13-MAY-1999 (first entry)
XX
DE Human PPAR-gamma-1 proximal promoter, exon A1 and intron A1.
XX
KW Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;
KW regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;
KW lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;
KW insulin resistance; non-insulin-dependent diabetes mellitus;
KW polycystic ovary syndrome; gastrointestinal tract; Crohn's disease;
KW inflammatory bowel disease; ulcerative colitis; bowel cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9905161-A1.
XX
PD 04-FEB-1999.
XX
PF 24-JUL-1998; 98WO-US015411.
XX
PR 25-JUL-1997; 97US-0053692P.
XX
PA (LIGA-) LIGAND PHARM INC.
PA (INSP) INST PASTEUR.
XX
PI Briggs MR, Saladin RS, Auwerx J, Fajas L;
XX
DR WPI; 1999-142844/12.
XX
PT Newly isolated nucleic acid comprising a control region of a human
PT peroxisome proliferator activated receptor (PPAR) gamma gene - useful for
PT identifying modulators that are useful in treating diseases associated
PT with abnormal levels of human PPAR-gamma gene expression.
XX
PS Claim 9; Page 78; 102pp; English.
XX
CC The present invention describes an isolated, purified or enriched nucleic
CC acid comprising a control region of a human peroxisome proliferator

CC activated receptor gamma (PPAR-gamma) gene. The nucleic acids are useful
CC for screening for agents capable of modulating the expression of a human
CC PPAR-gamma gene. These agents (modulators) form pharmaceutical
CC compositions that are useful for treating diseases associated with
CC high/low levels of human PPAR-gamma gene expression. The diseases include
CC obesity, anorexia, cachexia, lipodystrophy, lipomas, liposarcomas,
CC abnormalities associated with anti-human immunodeficiency virus (HIV)
CC treatment, insulin resistance, non-insulin-dependent diabetes mellitus
CC (NIDDM), polycystic ovary syndrome, diseases of the gastrointestinal (GI)
CC tract, inflammatory bowel disease, Crohn's disease, ulcerative colitis
CC and bowel cancer. The nucleic acids are useful for studying the role of
CC the PPAR-gamma gene in various diseases and disorders. The structure of
CC PPAR-gamma enables genetic studies of PPAR- gamma mutations in humans,
CC and evaluation of its role in disorders like insulin resistance, NIDDM,
CC and diseases associated with altered adipose tissue function, like
CC obesity and lipodystrophic syndromes. The nucleic acids are also useful
CC for gene therapy and the production of transgenic animals, which are
CC useful in screening assays. The control regions of the nucleic acids
CC enable screening for modulators of the human PPAR-gamma gene, which are
CC useful in designing drugs for treating disorders or diseases associated
CC with the level of PPAR-gamma gene expression. The present sequence
CC represents human PPAR-gamma-1 proximal promoter, exon A1 and intron A1
XX
SQ Sequence 503 BP; 60 A; 184 C; 211 G; 48 T; 0 U; 0 Other;

```
  Query Match             100.0%;  Score 125;  DB 2;  Length 503;
  Best Local Similarity   100.0%;  Pred. No. 7e-11;
  Matches  125;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCCAGCCGGCGCCCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCCAGCCGGCGCCCGC 60

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120

Qy        121 CCAGC 125
              |||||
Db        121 CCAGC 125
```

RESULT 2
ADS36487/c
ID ADS36487 standard; DNA; 12548 BP.
XX
AC ADS36487;
XX
DT 16-DEC-2004 (first entry)
XX
DE Human autoimmune disease-related genomic DNA sequence - SEQ ID 1701.
XX
KW single nucleotide polymorphism detection; SNP detection;
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW primary systemic vasculitis; ds.
XX
OS Homo sapiens.
XX
PN WO2004083403-A2.
XX
PD 30-SEP-2004.
XX
PF 18-MAR-2004; 2004WO-US008461.
XX
PR 18-MAR-2003; 2003US-0455444P.
PR 25-APR-2003; 2003US-0465241P.
XX
PA (APPL-) APPLERA CORP.
XX
PI Cargill M, Begovich AB, Alexander HC;
XX
DR WPI; 2004-728480/71.
XX
PT New isolated nucleic acid molecule comprises at least 8 contiguous
PT nucleotides where one of the nucleotides is a single nucleotide
PT polymorphism (SNP), useful for diagnosing or treating autoimmune
PT diseases, e.g. rheumatoid arthritis.
XX
PS Claim 16; SEQ ID NO 1701; 123pp; English.
XX
CC The invention comprises amino acid and coding sequences containing
CC genetic polymorphisms associated with an altered risk of developing an
CC autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC comprises a method of identifying an individual that has an altered risk
CC of developing an autoimmune disease, comprising detecting a single
CC nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC and protein sequences of the invention are useful for diagnosing and
CC treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC present nucleic acid represents a human autoimmune disease-related
CC genomic DNA sequence of the invention. NOTE: The present sequence is not
CC shown in the specification, but has been retrieved from the WIPO website.
XX
SQ Sequence 12548 BP; 3528 A; 2769 C; 2492 G; 3757 T; 0 U; 2 Other;

```
  Query Match             94.9%;  Score 118.6;  DB 13;  Length 12548;
  Best Local Similarity   96.8%;  Pred. No. 2.8e-10;
  Matches  121;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       7139 CCCCCGCCCCCGCCCCCGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 7080

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       7079 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGCTCCGCCGCGGGCAGGCGGGGC 7020

Qy        121 CCAGC 125
              |||||
Db       7019 CCAGC 7015
```

RESULT 3
ADS36461
ID ADS36461 standard; DNA; 158417 BP.
XX
AC ADS36461;
XX
DT 16-DEC-2004 (first entry)
XX
DE Human autoimmune disease-related genomic DNA sequence - SEQ ID 1675.
XX
KW single nucleotide polymorphism detection; SNP detection;
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW primary systemic vasculitis; ds.
XX
OS Homo sapiens.
XX
PN WO2004083403-A2.
XX
PD 30-SEP-2004.
XX
PF 18-MAR-2004; 2004WO-US008461.
XX
PR 18-MAR-2003; 2003US-0455444P.
PR 25-APR-2003; 2003US-0465241P.
XX

```
PA   (APPL-) APPLERA CORP.
XX
PI   Cargill M,  Begovich AB,  Alexander HC;
XX
DR   WPI; 2004-728480/71.
XX
PT   New isolated nucleic acid molecule comprises at least 8 contiguous
PT   nucleotides where one of the nucleotides is a single nucleotide
PT   polymorphism (SNP), useful for diagnosing or treating autoimmune
PT   diseases, e.g. rheumatoid arthritis.
XX
PS   Claim 16; SEQ ID NO 1675; 123pp; English.
XX
CC   The invention comprises amino acid and coding sequences containing
CC   genetic polymorphisms associated with an altered risk of developing an
CC   autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC   comprises a method of identifying an individual that has an altered risk
CC   of developing an autoimmune disease, comprising detecting a single
CC   nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC   and protein sequences of the invention are useful for diagnosing and
CC   treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC   diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC   bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC   anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC   myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC   present nucleic acid represents a human autoimmune disease-related
CC   genomic DNA sequence of the invention. NOTE: The present sequence is not
CC   shown in the specification, but has been retrieved from the WIPO website.
XX
SQ   Sequence 158417 BP; 47887 A; 30186 C; 31475 G; 48577 T; 0 U; 292 Other;

  Query Match             94.9%;  Score 118.6;  DB 13;  Length 158417;
  Best Local Similarity   96.8%;  Pred. No. 1.4e-10;
  Matches  121;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||
Db       5890 CCCCCGCCCCCGCCCCCGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 5949

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       5950 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGCTCCGCCGCGGGCAGGCGGGGC 6009

Qy        121 CCAGC 125
              |||||
Db       6010 CCAGC 6014


RESULT 4
AAX19063
ID   AAX19063 standard; DNA; 201 BP.
XX
AC   AAX19063;
XX
DT   13-MAY-1999  (first entry)
XX
DE   Human PPAR-gamma-1 proximal promoter.
XX
KW   Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;
KW   regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;
KW   lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;
KW   insulin resistance; non-insulin-dependent diabetes mellitus;
KW   polycystic ovary syndrome; gastrointestinal tract; Crohn's disease;
KW   inflammatory bowel disease; ulcerative colitis; bowel cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9905161-A1.
XX
PD   04-FEB-1999.
XX
PF   24-JUL-1998;   98WO-US015411.
XX
PR   25-JUL-1997;   97US-0053692P.
XX
PA   (LIGA-) LIGAND PHARM INC.
PA   (INSP ) INST PASTEUR.
XX
PI   Briggs MR,  Saladin RS,  Auwerx J,  Fajas L;
XX
DR   WPI; 1999-142844/12.
XX
PT   Newly isolated nucleic acid comprising a control region of a human
PT   peroxisome proliferator activated receptor (PPAR) gamma gene - useful for
PT   identifying modulators that are useful in treating diseases associated
PT   with abnormal levels of human PPAR-gamma gene expression.
XX
PS   Disclosure; Page 87; 102pp; English.
XX
CC   The present invention describes an isolated, purified or enriched nucleic
CC   acid comprising a control region of a human peroxisome proliferator
CC   activated receptor gamma (PPAR-gamma) gene. The nucleic acids are useful
CC   for screening for agents capable of modulating the expression of a human
CC   PPAR-gamma gene. These agents (modulators) form pharmaceutical
CC   compositions that are useful for treating diseases associated with
CC   high/low levels of human PPAR-gamma gene expression. The diseases include
CC   obesity, anorexia, cachexia, lipodystrophy, lipomas, liposarcomas,
CC   abnormalities associated with anti-human immunodeficiency virus (HIV)
CC   treatment, insulin resistance, non-insulin-dependent diabetes mellitus
CC   (NIDDM), polycystic ovary syndrome, diseases of the gastrointestinal (GI)
CC   tract, inflammatory bowel disease, Crohn's disease, ulcerative colitis
CC   and bowel cancer. The nucleic acids are useful for studying the role of
CC   the PPAR-gamma gene in various diseases and disorders. The structure of
CC   PPAR-gamma enables genetic studies of PPAR- gamma mutations in humans,
CC   and evaluation of its role in disorders like insulin resistance, NIDDM,
CC   and diseases associated with altered adipose tissue function, like
CC   obesity and lipodystrophic syndromes. The nucleic acids are also useful
CC   for gene therapy and the production of transgenic animals, which are
CC   useful in screening assays. The control regions of the nucleic acids
CC   enable screening for modulators of the human PPAR-gamma gene, which are
CC   useful in designing drugs for treating disorders or diseases associated
CC   with the level of PPAR-gamma gene expression. The present sequence
CC   represents the human PPAR-gamma-1 proximal promoter
XX
SQ   Sequence 201 BP; 23 A; 95 C; 73 G; 10 T; 0 U; 0 Other;

  Query Match             65.1%;  Score 81.4;  DB 2;  Length 201;
  Best Local Similarity   96.0%;  Pred. No. 0.00032;
  Matches   95;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    1;

Qy         30 ACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCCCGCCCCCGCGCCGGGCCCGGCTCGGC 89
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCCCGCCCCCGCGCCGGGCCCGGCTCGGC 60

Qy         90 CCGACCCGGTT---CCGCCGCGGGCAGGCGGGGCCCAGC 125
              ||||||||| |   |||||||||||||||||||||||||
Db         61 CCGACCCGGATCCGCCGCCGCGGGCAGGCGGGGCCCAGC 99


RESULT 5
ABZ20967/c
ID   ABZ20967 standard; DNA; 1416 BP.
XX
AC   ABZ20967;
XX
DT   29-APR-2003  (first entry)
XX
DE   Animal test kit marker DNA 127-sp6.
XX
KW   DNA marker; meat; consumer protection; genetic fingerprint; SNP;
KW   single nucleotide polymorphism; animal breeding; gene; ds.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(581,C)
```

```
FT                  /*tag= a
FT   variation       replace(697,A)
FT                  /*tag= b
XX
PN   DE20206477-U1.
XX
PD   10-OCT-2002.
XX
PF   14-MAR-2002; 2002DE-02006477.
XX
PR   14-MAR-2002; 2002DE-01012560.
PR   14-MAR-2002; 2002DE-02006477.
XX
PA   (GAGB-) GAG BIOSCIENCE GMBH.
XX
DR   WPI; 2003-048092/05.
XX
PT   Test kit containing bovine marker DNA, useful e.g. in animal breeding and
PT   for tracking foods, comprises specified polymorphic sequences.
XX
PS   Claim 1; Page 18; 25pp; German.
XX
CC   The present invention relates to a new test kit comprising marker DNA,
CC   for cattle or derived meat products, which has any of the sequences shown
CC   in ABZ20937-ABZ20970, all of which contain at least one variable position
CC   (single nucleotide polymorphism, SNP), and these SNPs are tabulated. The
CC   kit is used in animal production or breeding, and to track cattle-derived
CC   foods, e.g. for consumer protection. The specified sequences contain
CC   sufficient SNPs, present at various frequencies in the cattle population,
CC   to provide, in combination, a genetic fingerprint for individual animals.
CC   The variable positions can be queried by high-throughput methods, making
CC   possible typing of the entire cattle population in the form of a digital
CC   DNA signature. The present sequence is a bovine marker DNA for use in the
CC   test kit of the invention
XX
SQ   Sequence 1416 BP; 270 A; 253 C; 540 G; 348 T; 0 U; 5 Other;

  Query Match             46.2%;  Score 57.8;  DB 8;  Length 1416;
  Best Local Similarity   66.4%;  Pred. No. 0.7;
  Matches   83;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  ||||  ||||| ||||| ||||| ||||| |||||  ||  | | |||
Db        448 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCGC 389

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||  | |||||| | ||||| | |  ||  ||||   |||||| |    |   | |  |
Db        388 CCCCCGCGCCGCGCAGCGCCCGCCCCCCCCGCACCCCTCTCCGCCCCCCCCCCCCCCGCCC 329

Qy        121 CCAGC 125
              || ||
Db        328 CCCGC 324


RESULT 6
ACN51217/c
ID   ACN51217 standard; cDNA; 420 BP.
XX
AC   ACN51217;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Cotton androecium tissue EST Clone ID: LIB3828-011-Q1-N6-H6, SEQ:5998.
XX
KW   Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;
KW   variety Nucotton33B; library LIB3828; molecular tag; molecular marker;
KW   genetic mapping; molecular mapping; seed germination; plant growth;
KW   plant quality; plant yield; plant breeding; tissue printing; ss.
XX
OS   Gossypium hirsutum.
XX
PN   US2004123340-A1.
XX
PD   24-JUN-2004.
XX
PF   12-DEC-2001; 2001US-00021323.
XX
PR   14-DEC-2000; 2000US-0255619P.
XX
PA   (DEIK/) DEIKMAN J.
PA   (FENG/) FENG P C C.
PA   (FINC/) FINCHER K L.
PA   (ZIEG/) ZIEGLER T E.
XX
PI   Deikman J,  Feng PCC,  Fincher KL,  Ziegler TE;
XX
DR   WPI; 2004-479808/45.
XX
PT   New isolated nucleic acid molecule that encodes a plant protein or its
PT   fragment, useful for isolating a variety of agronomically significant
PT   genes associated with plant growth, quality or yield, and as molecular
PT   tags to map genes.
XX
PS   Claim 1; SEQ ID NO 5998; 34pp; English.
XX
CC   The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC   ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC   from primed or non-primed seeds from variety DP50B, mature seeds from
CC   variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC   tissue, developing fibres, carpel walls and septa from variety
CC   Nucotton33B. The invention also relates to substantially purified
CC   proteins or their fragments encoded by nucleic acid molecules of the
CC   invention, and to transformed plants having a nucleic acid construct
CC   comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC   molecular tags to isolate genetic regions, to isolate genes, to map
CC   genes, to determine gene function and to determining whether genes are
CC   members of a particular gene family. The nucleic acid molecules may be
CC   used for isolating a variety of agronomically significant genes
CC   associated with plant growth, quality, yield, and could also serve as
CC   links in metabolic and catabolic pathways. The nucleic acid molecules are
CC   also useful for identifying genes important in initiating and maintaining
CC   seed germination or that may be used to mitigate stresses encountered
CC   during seed germination. The ESTs additionally enable the acquisition of
CC   promoters and cis-regulatory elements which will be useful to express
CC   agronomically significant genes in these tissues and/or other tissues,
CC   and also permits the acquisition of molecular markers useful in breeding
CC   schemes, genetic and molecular mapping, and in cloning of agronomically
CC   significant genes. The nucleic acid molecules are further useful for
CC   detecting the expression level or pattern of a protein or mRNA and for
CC   detecting the presence or quantity of a protein by tissue printing. The
CC   present sequence represents a specifically claimed EST isolated from a
CC   cotton variety Nucotton33B androecium tissue cDNA library (LIB3828). The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the US
CC   patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
SQ   Sequence 420 BP; 8 A; 55 C; 342 G; 15 T; 0 U; 0 Other;

  Query Match             45.3%;  Score 56.6;  DB 13;  Length 420;
  Best Local Similarity   69.4%;  Pred. No. 1.5;
  Matches   77;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  ||||  ||||| ||||| ||||| ||||| |||||  ||  | ||| |
Db        145 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 86

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGC 111
              ||||||||| | |||  |||  | |  |||  ||| |  || || ||| |
Db         85 CCCCGCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCGGCCCCCCCCGGCC 35


RESULT 7
AAZ41312
ID   AAZ41312 standard; cDNA; 1315 BP.
XX
AC   AAZ41312;
```

XX
DT 18-JAN-2000 (first entry)
XX
DE Human normal ovarian tissue derived cDNA 91.
XX
KW Human; ovary; screening; ovarian cancer; treatment; ss.
XX
OS Homo sapiens.
XX
PN DE19816395-A1.
XX
PD 07-OCT-1999.
XX
PF 03-APR-1998; 98DE-01016395.
XX
PR 03-APR-1998; 98DE-01016395.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-552352/47.
XX
PT Nucleic acid sequences potentially useful in diagnosis or therapy of
PT ovarian cancer.
XX
PS Claim 3; Page 201; 274pp; German.
XX
CC This invention describes novel nucleic acid sequences that are highly
CC expressed in normal ovary tissue. Artificial chromosomes and cosmid
CC clones containing the sequences can be used as gene transfer vehicles.
CC The sequences can be used to produce DNA fragments containing full-length
CC genes. Host cells transformed with the sequences can be used to produce
CC polypeptides or polypeptide fragments, which can be used to screen phage
CC displays for polypeptides that bind to them, or as tools for identifying
CC agents active against ovarian cancer, or to prepare medicaments for
CC treating ovarian cancer. The cDNA sequences can be used to obtain genomic
CC genes, their promoters, enhancers, silencers, exon structures, intron
CC structures and their splice variants. AAZ41222-Z41324 represent cDNA
CC sequences derived from normal human ovarian tissue and which encode the
CC protein fragments represented in AAY59724-Y59837
XX
SQ Sequence 1315 BP; 355 A; 320 C; 228 G; 412 T; 0 U; 0 Other;

Query Match 43.5%; Score 54.4; DB 2; Length 1315;
Best Local Similarity 67.9%; Pred. No. 2.3;
Matches 76; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
Db 23 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGGCCGG 82

Qy 61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCA 112
Db 83 CCCCCCCCCCCCCCCCCCCCCCCCCTGACACGACATGGAACCCCGACCGGCA 134

RESULT 8
ABV61578
ID ABV61578 standard; cDNA; 200 BP.
XX
AC ABV61578;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 61569.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 11665-11666; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ Sequence 200 BP; 12 A; 106 C; 69 G; 9 T; 0 U; 4 Other;

Query Match 43.4%; Score 54.2; DB 5; Length 200;
Best Local Similarity 64.7%; Pred. No. 4.1;
Matches 77; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
Db 25 CCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCNCCACCCCCCNCCCCCCCCAGCCCACGC 84

Qy 61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGG 119
Db 85 CCCCGCCCCCACCCCCCCCACCCCCCCCCCCGNCCCCGCCANTGGGGGGGTTGGGGGGG 143

RESULT 9
ACN54594/c
ID ACN54594 standard; cDNA; 629 BP.
XX
AC ACN54594;
XX
DT 02-DEC-2004 (first entry)
XX
DE Cotton androecium tissue EST Clone ID: LIB3828-025-Q6-K6-D2, SEQ:9375.
XX
KW Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;
KW variety Nucotton33B; library LIB3828; molecular tag; molecular marker;
KW genetic mapping; molecular mapping; seed germination; plant growth;
KW plant quality; plant yield; plant breeding; tissue printing; ss.
XX
OS Gossypium hirsutum.
XX
PN US2004123340-A1.
XX
PD 24-JUN-2004.
XX

```
PF   12-DEC-2001; 2001US-00021323.
XX
PR   14-DEC-2000; 2000US-0255619P.
XX
PA   (DEIK/) DEIKMAN J.
PA   (FENG/) FENG P C C.
PA   (FINC/) FINCHER K L.
PA   (ZIEG/) ZIEGLER T E.
XX
PI   Deikman J,  Feng PCC,  Fincher KL,  Ziegler TE;
XX
DR   WPI; 2004-479808/45.
XX
PT   New isolated nucleic acid molecule that encodes a plant protein or its
PT   fragment, useful for isolating a variety of agronomically significant
PT   genes associated with plant growth, quality or yield, and as molecular
PT   tags to map genes.
XX
PS   Claim 1; SEQ ID NO 9375; 34pp; English.
XX
CC   The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC   ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC   from primed or non-primed seeds from variety DP50B, mature seeds from
CC   variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC   tissue, developing fibres, carpel walls and septa from variety
CC   Nucotton33B. The invention also relates to substantially purified
CC   proteins or their fragments encoded by nucleic acid molecules of the
CC   invention, and to transformed plants having a nucleic acid construct
CC   comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC   molecular tags to isolate genetic regions, to isolate genes, to map
CC   genes, to determine gene function and to determining whether genes are
CC   members of a particular gene family. The nucleic acid molecules may be
CC   used for isolating a variety of agronomically significant genes
CC   associated with plant growth, quality, yield, and could also serve as
CC   links in metabolic and catabolic pathways. The nucleic acid molecules are
CC   also useful for identifying genes important in initiating and maintaining
CC   seed germination or that may be used to mitigate stresses encountered
CC   during seed germination. The ESTs additionally enable the acquisition of
CC   promoters and cis-regulatory elements which will be useful to express
CC   agronomically significant genes in these tissues and/or other tissues,
CC   and also permits the acquisition of molecular markers useful in breeding
CC   schemes, genetic and molecular mapping, and in cloning of agronomically
CC   significant genes. The nucleic acid molecules are further useful for
CC   detecting the expression level or pattern of a protein or mRNA and for
CC   detecting the presence or quantity of a protein by tissue printing. The
CC   present sequence represents a specifically claimed EST isolated from a
CC   cotton variety Nucotton33B androecium tissue cDNA library (LIB3828). The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the US
CC   patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
SQ   Sequence 629 BP; 29 A; 38 C; 562 G; 0 T; 0 U; 0 Other;

  Query Match             43.4%;  Score 54.2;  DB 13;  Length 629;
  Best Local Similarity   69.2%;  Pred. No. 3;
  Matches   74;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||||||||  ||||  |||||  |||||  |||||  |||||  ||  ||   |||  ||||| |
Db        187 CCCCTGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCGTCGCCCCC 128

Qy         61 GCCCGCCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
                ||| |||||  | ||   |||  | |  |||  |||    || || |
Db        127 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCGCCCCCCC 81


RESULT 10
ADQ23807
ID   ADQ23807 standard; DNA; 1131 BP.
XX
AC   ADQ23807;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 6627.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 6627; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 1131 BP; 136 A; 416 C; 107 G; 119 T; 0 U; 353 Other;

  Query Match             42.4%;  Score 53;  DB 12;  Length 1131;
  Best Local Similarity   55.7%;  Pred. No. 3.9;
  Matches   68;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  | ||  ||||  ||||  ||| | ||||| |||||  ||||  ||  | ||  |
Db        119 CCCCCNCNCCCCCCCCCCCCCCCNCCCCNCNCCCCCCCCCCCCCNCCCCNNCCNNCCCCNNC 178

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
               ||| |||||   ||   |||  |    |||  |||    || || |   |       |
Db        179 CCCCCCCCCCCNNCCCNCCCCNNCCNNCCCCCCCCCCNNNCCCCCCCCCCCCNCNCNNNNCNC 238

Qy        121 CC 122
              ||
Db        239 CC 240


RESULT 11
ABN69641/c
ID   ABN69641 standard; DNA; 168 BP.
XX
AC   ABN69641;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 7195.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
```

KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus agalactiae.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB004789.
XX
PR 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
PI Tettelin H;
XX
DR WPI; 2002-352536/38.
DR P-PSDB; ABP29010.
XX
PT New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
PS Claim 7; Page 3875; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyrogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 168 BP; 4 A; 4 C; 147 G; 7 T; 0 U; 6 Other;

Query Match 41.8%; Score 52.2; DB 6; Length 168;
Best Local Similarity 67.3%; Pred. No. 8.5;
Matches 72; Conservative 1; Mismatches 34; Indels 0; Gaps 0;

```
Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  |||| |||||| ||||| ||||| ||||| |||||  :|  | |||||
Db        117 CCCCCCCCCCCCCCCCGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCMCCCCCCCGC 58

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
                 ||| ||||| | ||   |||  | |  |||  |||    || || |
Db         57 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 11
```

RESULT 12
ADC87060/c
ID ADC87060 standard; DNA; 3163 BP.
XX
AC ADC87060;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:1513.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87061.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 1513; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 3163 BP; 340 A; 345 C; 1360 G; 610 T; 0 U; 508 Other;

Query Match 41.6%; Score 52; DB 10; Length 3163;
Best Local Similarity 60.8%; Pred. No. 4.3;
Matches 76; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  |||| |||||| ||||| ||||| ||||| |||||  ||  | ||| |
Db        872 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCNCCCCC 813

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
                ||| ||||| | ||   |||  | |  |||  |||    || || |   |   |    |
Db        812 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCNCCCCNNNC 753

Qy        121 CCAGC 125
                | |
Db        752 NCNCC 748
```

RESULT 13
ADG37080
ID ADG37080 standard; DNA; 28198 BP.
XX
AC ADG37080;
XX
DT 26-FEB-2004 (first entry)
XX
DE Mouse plakoglobin polynucleotide seq id 2.
XX
KW animal cell; beta-catenin; plakoglobin; cadherin adhesion;
KW signal transduction; cross-reactivity; desmosome cadherin;
KW Wnt signal transduction; mouse; ds.
XX
OS Mus musculus.
XX
PN JP2003189850-A.
XX

PD 08-JUL-2003.
XX
PF 27-DEC-2001; 2001JP-00398640.
XX
PR 27-DEC-2001; 2001JP-00398640.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR WPI; 2003-819446/77.
XX
PT New animal cell defective in or lacking expression of beta-catenin or
PT plakoglobin, is useful for studying the function and regulation mechanism
PT beta-catenin and/or plakoglobin, and in developing therapeutic agents.
XX
PS Disclosure; SEQ ID NO 2; 41pp; Japanese.
XX
CC The invention describes an animal cell which has defective or lack of
CC expression of beta-catenin and/or plakoglobin. The cell is useful for
CC research purposes. The beta-catenin defective cell, plakoglobin defective
CC cell or the beta-catenin and plakoglobin double defective animal cells
CC are useful for studying the function and regulation mechanism beta-
CC catenin and/or plakoglobin, for analysing the diseases, disorders or
CC conditions resulting from the variation/mutation of the proteins,
CC developing therapeutic agents and improving treatment methods. The cells
CC are useful for (a) elucidating cadherin adhesion by beta-catenin and
CC plakoglobin, (b) elucidating signal transduction mechanism for cell
CC adhesion through beta-catenin and plakoglobin, (c) elucidating cross-
CC reactivity of cadherin and desmosome cadherin, (d) elucidating function
CC of beta-catenin in Wnt signal transduction, (e) elucidating signal
CC transduction of plakoglobin. The cells enable effective analysis of the
CC beta-catenin and plakoglobin functions in vitro. This sequence represents
CC a mouse plakoglobin polynucleotide.
XX
SQ Sequence 28198 BP; 6568 A; 7442 C; 7158 G; 7019 T; 0 U; 11 Other;

Query Match 41.6%; Score 52; DB 10; Length 28198;
Best Local Similarity 63.2%; Pred. No. 2.4;
Matches 79; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
Db 13859 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 13918

Qy 61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
Db 13919 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNAGTT 13978

Qy 121 CCAGC 125
Db 13979 CCAGC 13983

RESULT 14
AAI82007/c
ID AAI82007 standard; cDNA; 344 BP.
XX
AC AAI82007;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2067.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO02076.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 1; SEQ ID NO 2067; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 344 BP; 33 A; 26 C; 225 G; 40 T; 0 U; 20 Other;

Query Match 41.4%; Score 51.8; DB 4; Length 344;
Best Local Similarity 61.2%; Pred. No. 8.1;
Matches 74; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
Db 286 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 227

Qy 61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
Db 226 CCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCNCCNNNNNNNC 167

Qy 121 C 121
Db 166 C 166

RESULT 15
ACN58918/c
ID ACN58918 standard; cDNA; 437 BP.
XX
AC ACN58918;
XX
DT 02-DEC-2004 (first entry)
XX
DE Cotton gynoecium tissue EST Clone ID: LIB3829-013-Q6-N6-G12, SEQ:13699.
XX
KW Cotton; plant; EST; expressed sequence tag; transgenic plant; gynoecium;
KW variety Nucotton33B; library LIB3829; molecular tag; molecular marker;
KW genetic mapping; molecular mapping; seed germination; plant growth;
KW plant quality; plant yield; plant breeding; tissue printing; ss.
XX
OS Gossypium hirsutum.
XX
PN US2004123340-A1.
XX
PD 24-JUN-2004.
XX
PF 12-DEC-2001; 2001US-00021323.
XX

PR 14-DEC-2000; 2000US-0255619P.
XX
PA (DEIK/) DEIKMAN J.
PA (FENG/) FENG P C C.
PA (FINC/) FINCHER K L.
PA (ZIEG/) ZIEGLER T E.
XX
PI Deikman J, Feng PCC, Fincher KL, Ziegler TE;
XX
DR WPI; 2004-479808/45.
XX
PT New isolated nucleic acid molecule that encodes a plant protein or its
PT fragment, useful for isolating a variety of agronomically significant
PT genes associated with plant growth, quality or yield, and as molecular
PT tags to map genes.
XX
PS Claim 1; SEQ ID NO 13699; 34pp; English.
XX
CC The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC from primed or non-primed seeds from variety DP50B, mature seeds from
CC variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC tissue, developing fibres, carpel walls and septa from variety
CC Nucotton33B. The invention also relates to substantially purified
CC proteins or their fragments encoded by nucleic acid molecules of the
CC invention, and to transformed plants having a nucleic acid construct
CC comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC molecular tags to isolate genetic regions, to isolate genes, to map
CC genes, to determine gene function and to determining whether genes are
CC members of a particular gene family. The nucleic acid molecules may be
CC used for isolating a variety of agronomically significant genes
CC associated with plant growth, quality, yield, and could also serve as
CC links in metabolic and catabolic pathways. The nucleic acid molecules are
CC also useful for identifying genes important in initiating and maintaining
CC seed germination or that may be used to mitigate stresses encountered
CC during seed germination. The ESTs additionally enable the acquisition of
CC promoters and cis-regulatory elements which will be useful to express
CC agronomically significant genes in these tissues and/or other tissues,
CC and also permits the acquisition of molecular markers useful in breeding
CC schemes, genetic and molecular mapping, and in cloning of agronomically
CC significant genes. The nucleic acid molecules are further useful for
CC detecting the expression level or pattern of a protein or mRNA and for
CC detecting the presence or quantity of a protein by tissue printing. The
CC present sequence represents a specifically claimed EST isolated from a
CC cotton variety Nucotton33B gynoecium tissue cDNA library (LIB3829). The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the US
CC patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
SQ Sequence 437 BP; 55 A; 50 C; 300 G; 32 T; 0 U; 0 Other;

Query Match 41.4%; Score 51.8; DB 13; Length 437;
Best Local Similarity 66.7%; Pred. No. 7.6;
Matches 74; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| | ||   ||||| ||||| ||||| ||||| ||||| |||||  ||  | |||
Db        362 CCCCCGGCCAACCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCTG 303

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGC 111
                |   || || | || ||||||||  |||| | |   |  || || | |||
Db        302 CCTTCCCGCCCCTCCCGGCCCGGCCTGGCCGGCCTGCGGCCCCCCCCCGGC 252
```

Search completed: November 9, 2005, 00:12:35
Job time : 104.246 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 15:16:42 ; Search time 31.8505 Seconds
(without alignments)
6421.701 Million cell updates/sec

Title: US-09-463-542-1_COPY_1_125
Perfect score: 125
Sequence: 1 cccctgcccctgcccctgcc.....gcgggcaggcggggcccagc 125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	52.6	42.1	865	4	US-09-270-767-11042	Sequence 11042, A
c	2	51.4	41.1	670689	4	US-09-949-016-12505	Sequence 12505, A
c	3	51.4	41.1	670690	4	US-09-949-016-14207	Sequence 14207, A
	4	50.6	40.5	152331	3	US-09-128-155-16	Sequence 16, Appl
	5	50.2	40.2	1845	4	US-09-614-034-188	Sequence 188, App
c	6	49.8	39.8	204	4	US-09-107-433-2184	Sequence 2184, Ap
	7	49.8	39.8	209	4	US-09-107-433-185	Sequence 185, App
	8	49.8	39.8	282	4	US-09-107-433-184	Sequence 184, App
c	9	49.8	39.8	308	4	US-09-107-433-1723	Sequence 1723, Ap
c	10	49.8	39.8	612	4	US-09-107-433-1970	Sequence 1970, Ap
c	11	49.6	39.7	320	3	US-09-165-264-11	Sequence 11, Appl
c	12	49.4	39.5	318	3	US-09-165-264-12	Sequence 12, Appl
c	13	49.4	39.5	319	3	US-09-165-264-8	Sequence 8, Appli
c	14	49.4	39.5	320	3	US-09-165-264-7	Sequence 7, Appli
c	15	49.4	39.5	320	3	US-09-165-264-13	Sequence 13, Appl
c	16	49.4	39.5	320	3	US-09-165-264-14	Sequence 14, Appl
	17	49	39.2	11766	4	US-09-949-016-12531	Sequence 12531, A
	18	49	39.2	11770	4	US-09-949-016-12720	Sequence 12720, A
	19	49	39.2	11770	4	US-09-949-016-13487	Sequence 13487, A
	20	49	39.2	11770	4	US-09-949-016-13488	Sequence 13488, A
c	21	48.8	39.0	22206	4	US-09-949-016-13901	Sequence 13901, A
c	22	47	37.6	336	4	US-09-543-681A-4150	Sequence 4150, Ap
c	23	47	37.6	339	4	US-09-543-681A-4147	Sequence 4147, Ap
c	24	47	37.6	483	4	US-09-543-681A-4138	Sequence 4138, Ap
	25	47	37.6	516	4	US-09-543-681A-4140	Sequence 4140, Ap
	26	47	37.6	519	4	US-09-543-681A-4145	Sequence 4145, Ap
	27	47	37.6	608	4	US-09-543-681A-4115	Sequence 4115, Ap
c	28	46.8	37.4	601	4	US-09-949-016-24803	Sequence 24803, A
c	29	46.8	37.4	601	4	US-09-949-016-24804	Sequence 24804, A
c	30	46.8	37.4	601	4	US-09-949-016-67084	Sequence 67084, A
c	31	46.8	37.4	601	4	US-09-949-016-67085	Sequence 67085, A
c	32	46.8	37.4	25633	4	US-09-949-016-12084	Sequence 12084, A
c	33	46.8	37.4	25633	4	US-09-949-016-13721	Sequence 13721, A
c	34	46.6	37.3	31199	4	US-09-949-016-16516	Sequence 16516, A
	35	46.4	37.1	1683	4	US-09-949-016-1977	Sequence 1977, Ap
c	36	46.4	37.1	9364	4	US-09-949-016-16777	Sequence 16777, A
	37	46.4	37.1	14688	4	US-09-949-016-12220	Sequence 12220, A
	38	46.4	37.1	14797	4	US-09-949-016-13719	Sequence 13719, A
	39	46.4	37.1	22976	4	US-09-269-939A-19	Sequence 19, Appl
	40	46.2	37.0	22431	4	US-09-949-016-14099	Sequence 14099, A
c	41	46	36.8	40546	4	US-09-949-016-12847	Sequence 12847, A
c	42	46	36.8	40546	4	US-09-949-016-12915	Sequence 12915, A
c	43	45.8	36.6	492	4	US-09-854-133-323	Sequence 323, App
	44	45.2	36.2	12425	4	US-09-616-289-50	Sequence 50, Appl
	45	45.2	36.2	31467	4	US-09-949-016-13134	Sequence 13134, A

ALIGNMENTS

RESULT 1
US-09-270-767-11042
; Sequence 11042, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11042
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: n means any nucleotide
US-09-270-767-11042

Query Match 42.1%; Score 52.6; DB 4; Length 865;
Best Local Similarity 68.2%; Pred. No. 0.1;
Matches 73; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| |||| ||||| ||||| || || ||||| |||||  ||  | ||| |
Db        159 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCC 218

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
                ||| ||||||| ||   ||| || |  |||  |||    || ||||
Db        219 CCCCCCCCCCGCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCGC 265
```

RESULT 2
US-09-949-016-12505/c
; Sequence 12505, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08

```
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12505
;   LENGTH: 670689
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(670689)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12505

  Query Match             41.1%;  Score 51.4;  DB 4;  Length 670689;
  Best Local Similarity   67.0%;  Pred. No. 0.071;
  Matches   73;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          7 CCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCCCGC 66
              |||| ||| |||||| ||||| |||| ||| ||| |||| | |||||| ||||| |
Db      40266 CCCCGACCCGCGCCCCCGCCCCCCGCCCCCCCCCGCACGCCCACGCCGCGCCCCCGCCCCC 40207

Qy         67 CCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGC 115
              |||| | ||   ||| | |  ||| |||||  || |  |  || | |
Db      40206 CCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCGGGCCCCCGACCCGCGGCC 40158


RESULT 3
US-09-949-016-14207/c
; Sequence 14207, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14207
;   LENGTH: 670690
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(670690)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14207

  Query Match             41.1%;  Score 51.4;  DB 4;  Length 670690;
  Best Local Similarity   67.0%;  Pred. No. 0.071;
  Matches   73;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          7 CCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCCCGC 66
              |||| ||| |||||| ||||| |||| ||| ||| |||| | |||||| ||||| |
Db      40266 CCCCGACCCGCGCCCCCGCCCCCCGCCCCCCCCCGCACGCCCACGCCGCGCCCCCGCCCCC 40207

Qy         67 CCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGC 115
              |||| | ||   ||| | |  ||| |||||  || |  |  || | |
Db      40206 CCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCGGGCCCCCGACCCGCGGCC 40158


RESULT 4
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
;   LENGTH: 152331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(152331)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

  Query Match             40.5%;  Score 50.6;  DB 3;  Length 152331;
  Best Local Similarity   66.4%;  Pred. No. 0.12;
  Matches   71;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              | ||  ||||  ||||  ||||| ||||| ||||| || || |   | | | ||||| |
Db      21876 CNCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCGCCCCGNGCGGGCCGCGCCCCC 21935

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
              ||||||||| || |  ||||  | ||||||  | |    || || |
Db      21936 CCCCGCCCCCCCGGCCCGCCCCCCGCGGCCCCCCACCCCCCCCCCC 21982


RESULT 5
US-09-614-034-188
; Sequence 188, Application US/09614034
; Patent No. 6489307
; GENERAL INFORMATION:
;  APPLICANT: PHILLIPS, M. IAN
;  APPLICANT:  ZHANG, YUAN
;  TITLE OF INVENTION: ANTISENSE COMPOSITIONS TARGETED TO BETA1-ADRENOCEPTOR-SPECIFIC mRI
;  TITLE OF INVENTION:  METHODS OF USE
;  FILE REFERENCE: 4300.013900
;  CURRENT APPLICATION NUMBER: US/09/614,034
;  CURRENT FILING DATE:  2000-07-11
;  PRIOR APPLICATION NUMBER: 09/152,717
;  PRIOR FILING DATE: 1998-09-14
;  PRIOR APPLICATION NUMBER: PCT/US99/21007
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 204
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 188
;   LENGTH: 1845
;   TYPE: DNA
;   ORGANISM: UNKNOWN
;   FEATURE:
;   OTHER INFORMATION: SYNTHETIC OLIGONUCLEOTIDE
;   NAME/KEY: misc_feature
;   LOCATION: (1828)..(1828)
;   OTHER INFORMATION: WHERE N = A, T, C OR G
US-09-614-034-188

  Query Match             40.2%;  Score 50.2;  DB 4;  Length 1845;
  Best Local Similarity   68.0%;  Pred. No. 0.25;
  Matches   70;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         16 CTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCCCGCCCCCGCGCC 75
              | |||||   |  | ||||  ||||  |||  ||| ||||||||||||| | ||||||||
Db       1130 CGGCCCCGCGCGGCCCCCCGCGCCCCCGCCCGCGCCCGCGCCCGCGCCCCCGCCCGCGCC 1189

Qy         76 GGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGG 118
                 ||| |  | || |||| | |||   | |||| |||   ||||
```

```
Db       1190 CGGCTCCCCGCGCCCCGCCGCGGCCGCCCCGCTGGCCAACGGG 1232


RESULT 6
US-09-107-433-2184/c
; Sequence 2184, Application US/09107433
; Patent No. 6800744
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                           SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNO
;                           THERAPEUTICS
;       NUMBER OF SEQUENCES: 5206
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: <Unknown>
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: <Unknown>
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,433
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/ 085131
;            FILING DATE: May 12, 1998
;            APPLICATION NUMBER: 60/051553
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-011
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;  INFORMATION FOR SEQ ID NO: 2184:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 204 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: circular
;       MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Streptococcus pneumoniae
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...204
;       SEQUENCE DESCRIPTION: SEQ ID NO: 2184:
US-09-107-433-2184

  Query Match             39.8%;  Score 49.8;  DB 4;  Length 204;
  Best Local Similarity   59.0%;  Pred. No. 0.39;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| ||||  | ||  |||| ||||| ||||| ||||| |||||  || | | | |
Db        203 CCCCCCCCCCCCCCNCCCNCCCCNCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCNNCCCNCCC 144

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||| ||||| | ||   ||| | |  ||| |||    || || |       |   |
Db        143 CCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCNNCNCCCCNCNCC 84

Qy        121 CC 122
              ||
Db         83 CC 82


RESULT 7
US-09-107-433-185
; Sequence 185, Application US/09107433
; Patent No. 6800744
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                           SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNO
;                           THERAPEUTICS
;       NUMBER OF SEQUENCES: 5206
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: <Unknown>
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: <Unknown>
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,433
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/ 085131
;            FILING DATE: May 12, 1998
;            APPLICATION NUMBER: 60/051553
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-011
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;  INFORMATION FOR SEQ ID NO: 185:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 209 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: circular
;       MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Streptococcus pneumoniae
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...209
;       SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-09-107-433-185

  Query Match             39.8%;  Score 49.8;  DB 4;  Length 209;
  Best Local Similarity   59.0%;  Pred. No. 0.39;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| ||||  | ||  |||| ||||| ||||| ||||| |||||  || | | | |
Db          3 CCCCCCCCCCCCCCNCCCNCCCCNCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCNNCCCNCCC 62

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||| ||||| | ||   ||| | |  ||| |||    || || |       |   |
Db         63 CCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCNNCNCCCCNCNCC 122

Qy        121 CC 122
              ||
Db        123 CC 124


RESULT 8
```

```
US-09-107-433-184
; Sequence 184, Application US/09107433
; Patent No. 6800744
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                            SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNO
;                            THERAPEUTICS
;        NUMBER OF SEQUENCES: 5206
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: <Unknown>
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: <Unknown>
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,433
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/ 085131
;             FILING DATE: May 12, 1998
;             APPLICATION NUMBER: 60/051553
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-011
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;    INFORMATION FOR SEQ ID NO: 184:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 282 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Streptococcus pneumoniae
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...282
;        SEQUENCE DESCRIPTION: SEQ ID NO: 184:
US-09-107-433-184

  Query Match             39.8%;  Score 49.8;  DB 4;  Length 282;
  Best Local Similarity   59.0%;  Pred. No. 0.37;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  | ||  ||||  ||||| ||||| ||||| |||||  ||  | | | |
Db          2 CCCCCCCCCCCCCNCCCNCCCCNCCCCCCCCCCCCNCCCCCCCCCCCCCCCNNCCCNCCC 61

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
                ||| ||||| | ||   |||  | |  |||  |||    || || |      |    |
Db         62 CCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCNNCNCCCCNCNCC 121

Qy        121 CC 122
              ||
Db        122 CC 123


RESULT 9
US-09-107-433-1723/c
; Sequence 1723, Application US/09107433
; Patent No. 6800744
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                            SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNO
;                            THERAPEUTICS
;        NUMBER OF SEQUENCES: 5206
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: <Unknown>
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: <Unknown>
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,433
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/ 085131
;             FILING DATE: May 12, 1998
;             APPLICATION NUMBER: 60/051553
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-011
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;    INFORMATION FOR SEQ ID NO: 1723:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 308 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        ORIGINAL SOURCE:
;             ORGANISM: Streptococcus pneumoniae
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...308
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1723:
US-09-107-433-1723

  Query Match             39.8%;  Score 49.8;  DB 4;  Length 308;
  Best Local Similarity   59.0%;  Pred. No. 0.37;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  | ||  ||||  ||||| ||||| ||||| |||||  ||  | | | |
Db        306 CCCCCCCCCCCCCNCCCNCCCCNCCCCCCCCCCCCNCCCCCCCCCCCCCCCNNCCCNCCC 247

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
                ||| ||||| | ||   |||  | |  |||  |||    || || |      |    |
Db        246 CCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCNNCNCCCCNCNCC 187

Qy        121 CC 122
              ||
Db        186 CC 185


RESULT 10
US-09-107-433-1970/c
; Sequence 1970, Application US/09107433
; Patent No. 6800744
;   GENERAL INFORMATION:
```

```
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
;                           SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE   FOR DIAGNO
;                           THERAPEUTICS
;       NUMBER OF SEQUENCES: 5206
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: <Unknown>
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: <Unknown>
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,433
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/ 085131
;            FILING DATE: May 12, 1998
;            APPLICATION NUMBER: 60/051553
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-011
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 1970:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 612 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: circular
;       MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Streptococcus pneumoniae
;       FEATURE:
;            NAME/KEY:  misc_feature
;            LOCATION: (B) LOCATION 1...612
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1970:
US-09-107-433-1970

  Query Match             39.8%;  Score 49.8;  DB 4;  Length 612;
  Best Local Similarity   59.0%;  Pred. No. 0.34;
  Matches   72;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| | ||  |||| ||||| ||||| ||||| |||||  ||  | | | |
Db        610 CCCCCCCCCCCCCCNCCCNCCCCNCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCNNCCCNCCC 551

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
                ||| ||||| | ||   |||  | |  |||  |||     || || |       |   |
Db        550 CCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCNNCNCCCCNCNCC 491

Qy        121 CC 122
              ||
Db        490 CC 489


RESULT 11
US-09-165-264-11/c
; Sequence 11, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 11
;   LENGTH: 320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-11

  Query Match             39.7%;  Score 49.6;  DB 3;  Length 320;
  Best Local Similarity   67.3%;  Pred. No. 0.4;
  Matches   70;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          4 CTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCC 63
              |||||||  ||||  ||||| ||||| ||||| ||||| |||||  ||  | ||| | ||
Db        304 CTGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 245

Qy         64 CGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
              | ||||| | ||   |||  | |  |||  |||    || || |
Db        244 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 201


RESULT 12
US-09-165-264-12/c
; Sequence 12, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 318
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12

  Query Match             39.5%;  Score 49.4;  DB 3;  Length 318;
  Best Local Similarity   66.4%;  Pred. No. 0.43;
  Matches   71;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| |||| ||||| ||||| ||||| ||||| |||||  ||  | ||| |
Db        300 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 241

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
               ||| ||||| | ||   |||  | |  |||  |||    || || |
Db        240 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 194


RESULT 13
US-09-165-264-8/c
; Sequence 8, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 319
```

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-8

  Query Match             39.5%;  Score 49.4;  DB 3;  Length 319;
  Best Local Similarity   66.4%;  Pred. No. 0.43;
  Matches   71;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| |||| ||||| ||||| ||||| ||||| |||||  ||  | ||| |
Db        300 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 241

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
              ||| ||||| | ||   |||  | |  |||  |||    || || |
Db        240 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 194


RESULT 14
US-09-165-264-7/c
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

  Query Match             39.5%;  Score 49.4;  DB 3;  Length 320;
  Best Local Similarity   66.4%;  Pred. No. 0.43;
  Matches   71;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| |||| ||||| ||||| ||||| ||||| |||||  ||  | ||| |
Db        300 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 241

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
              ||| ||||| | ||   |||  | |  |||  |||    || || |
Db        240 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 194


RESULT 15
US-09-165-264-13/c
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13

  Query Match             39.5%;  Score 49.4;  DB 3;  Length 320;
  Best Local Similarity   66.4%;  Pred. No. 0.43;
  Matches   71;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| |||| ||||| ||||| ||||| ||||| |||||  ||  | ||| |
Db        300 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 241

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
              ||| ||||| | ||   |||  | |  |||  |||    || || |
Db        240 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 194


Search completed: November  8, 2005, 19:52:44
Job time : 33.8505 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 18:06:53 ; Search time 161.744 Seconds
(without alignments)
6391.142 Million cell updates/sec

Title: US-09-463-542-1_COPY_1_125
Perfect score: 125
Sequence: 1 cccctgcccctgcccctgcc.....gcgggcaggcggggcccagc 125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	57.8	46.2	667	20	US-10-767-701-8805	Sequence 8805, Ap
	2	57.6	46.1	737	21	US-10-425-115-65924	Sequence 65924, A
c	3	56.8	45.4	972	21	US-10-425-115-17653	Sequence 17653, A
c	4	56.6	45.3	420	20	US-10-021-323-5998	Sequence 5998, Ap
	5	56.2	45.0	668	20	US-10-437-963-3757	Sequence 3757, Ap
	6	56	44.8	678	20	US-10-767-701-8435	Sequence 8435, Ap
	7	55.8	44.6	766	21	US-10-425-115-82176	Sequence 82176, A
c	8	55.6	44.5	809	19	US-10-424-599-26988	Sequence 26988, A
	9	55.4	44.3	862	20	US-10-437-963-28390	Sequence 28390, A
	10	55.2	44.2	956	14	US-10-027-632-31508	Sequence 31508, A
	11	55.2	44.2	956	18	US-10-027-632-31508	Sequence 31508, A
	12	55	44.0	713	20	US-10-767-701-5654	Sequence 5654, Ap
	13	54.8	43.8	581	21	US-10-425-115-93481	Sequence 93481, A
	14	54.8	43.8	788	21	US-10-425-115-63819	Sequence 63819, A
	15	54.8	43.8	793	20	US-10-767-701-9291	Sequence 9291, Ap
c	16	54.6	43.7	358	19	US-10-424-599-99880	Sequence 99880, A
	17	54.6	43.7	1470	20	US-10-767-701-14291	Sequence 14291, A
	18	54.6	43.7	2432	20	US-10-437-963-64501	Sequence 64501, A
c	19	54.4	43.5	587	21	US-10-425-115-91308	Sequence 91308, A
	20	54.2	43.4	200	21	US-10-357-930-61597	Sequence 61597, A
c	21	54.2	43.4	629	20	US-10-021-323-9375	Sequence 9375, Ap
	22	54.2	43.4	891	21	US-10-425-115-146045	Sequence 146045,
c	23	54	43.2	805	19	US-10-424-599-61903	Sequence 61903, A
	24	53.6	42.9	673	21	US-10-425-115-11145	Sequence 11145, A
c	25	53.6	42.9	721	19	US-10-424-599-139811	Sequence 139811,
	26	53.6	42.9	728	21	US-10-425-115-183861	Sequence 183861,
	27	53.4	42.7	658	20	US-10-767-701-4204	Sequence 4204, Ap
	28	53.4	42.7	700	21	US-10-425-115-14077	Sequence 14077, A
c	29	53.2	42.6	448	21	US-10-425-115-108351	Sequence 108351,
	30	53.2	42.6	712	21	US-10-425-115-118813	Sequence 118813,
	31	53	42.4	706	20	US-10-437-963-17102	Sequence 17102, A
	32	53	42.4	763	20	US-10-437-963-49370	Sequence 49370, A
	33	53	42.4	1131	21	US-10-723-860-6627	Sequence 6627, Ap
	34	53	42.4	1215	20	US-10-767-701-8767	Sequence 8767, Ap
c	35	52.8	42.2	590	21	US-10-425-115-133228	Sequence 133228,
	36	52.8	42.2	821	21	US-10-425-115-131043	Sequence 131043,
	37	52.6	42.1	492	19	US-10-424-599-59619	Sequence 59619, A
	38	52.6	42.1	497	19	US-10-424-599-133788	Sequence 133788,
	39	52.6	42.1	639	20	US-10-767-701-5995	Sequence 5995, Ap
c	40	52.6	42.1	740	21	US-10-425-115-131885	Sequence 131885,
c	41	52.6	42.1	985	21	US-10-425-115-91623	Sequence 91623, A
	42	52.4	41.9	440	21	US-10-425-115-106153	Sequence 106153,
	43	52.4	41.9	822	21	US-10-425-115-102047	Sequence 102047,
c	44	52.2	41.8	771	19	US-10-424-599-107958	Sequence 107958,
	45	52.2	41.8	1533	21	US-10-739-930-4713	Sequence 4713, Ap

ALIGNMENTS

RESULT 1
US-10-767-701-8805
; Sequence 8805, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 8805
; LENGTH: 667
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS56975_1
US-10-767-701-8805

Query Match 46.2%; Score 57.8; DB 20; Length 667;
Best Local Similarity 66.4%; Pred. No. 6.7e-05;
Matches 83; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60

```
Db        182 CCCCCGCCCCCGCCCCTCCCCCCCCCCCCCCCCCACCTCCCCCCCCGCCGCGCCCTCCCC 241

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||| ||||| |||||  || | |  | | ||  |||   || || | | |  |||||||
Db        242 CCCCCCCCCCCCGCCGCCCCGGCCCGGCCGCGTGCCGCCCCCCCCCCCGCCGCGCGGGGC 301

Qy        121 CCAGC 125
                ||
Db        302 GGGGC 306


RESULT 2
US-10-425-115-65924
; Sequence 65924, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 65924
;   LENGTH: 737
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(737)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_160123C.1
US-10-425-115-65924

  Query Match             46.1%;  Score 57.6;  DB 21;  Length 737;
  Best Local Similarity   61.5%;  Pred. No. 7.1e-05;
  Matches   75;  Conservative    9;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  :|||  :||||:||||| ||||| ||||| |||||  ||  ||||| |
Db        558 CCCCCCCCCCCCCMCCCMCMCCCCMCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCGCCCCC 617

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||||||||| | |:   ||:  | |   ||  ||||    | || ::||| |:| | :
Db        618 CCCCGCCCCCCCCCCMCCCCCMYCCVCACACCCSCCCGWACGCCCCCSKGGCCGSCCGCKA 677

Qy        121 CC 122
              ||
Db        678 CC 679


RESULT 3
US-10-425-115-17653/c
; Sequence 17653, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 17653
;   LENGTH: 972
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(972)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_116102C.1
US-10-425-115-17653

  Query Match             45.4%;  Score 56.8;  DB 21;  Length 972;
  Best Local Similarity   66.1%;  Pred. No. 9.8e-05;
  Matches   82;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  || | |||||  ||||| || || |||||  || | | ||  |||  | ||  |
Db        707 CCCCCCCCGCCGCCCCGCCCCCCCCCGCCCCCCCCCGCCGCCCACCGCGCCCCCCCCGCC 648

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||| ||||| | |  | |||| | || ||||||||||  ||  |||| |   ||| ||
Db        647 CCCCCCCCCCCCCCGCGCCCCGCCCCGCCCCGACCCGGCGCCCGCGCGCGAGCGCGCGGG 588

Qy        121 CCAG 124
              | ||
Db        587 CGAG 584


RESULT 4
US-10-021-323-5998/c
; Sequence 5998, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
;  APPLICANT: Deikman, Jill
;  APPLICANT:  Feng, Paul C.C.
;  APPLICANT:  Fincher, Karen L.
;  APPLICANT:  Ziegler, Todd E.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(52274)B
;  CURRENT APPLICATION NUMBER: US/10/021,323
;  CURRENT FILING DATE:  2001-12-12
;  PRIOR APPLICATION NUMBER: US 60/255, 619
;  PRIOR FILING DATE: 2000-12-14
;  NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 5998
;   LENGTH: 420
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3828-011-Q1-N6-H6
US-10-021-323-5998

  Query Match             45.3%;  Score 56.6;  DB 20;  Length 420;
  Best Local Similarity   69.4%;  Pred. No. 0.00015;
  Matches   77;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  ||||  ||||| ||||| ||||| ||||| |||||  ||  | ||| |
Db        145 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 86

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGC 111
              ||||||||| | |||  |||  | |  |||  ||| |  || || ||| |
Db         85 CCCCGCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCGGCCCCCCCCGGCC 35


RESULT 5
US-10-437-963-3757
; Sequence 3757, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
```

```
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 3757
;   LENGTH: 668
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(668)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_10702C.1
US-10-437-963-3757

  Query Match             45.0%;  Score 56.2;  DB 20;  Length 668;
  Best Local Similarity   65.6%;  Pred. No. 0.00015;
  Matches   82;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| |||| ||||| |||| ||||| ||||| |||||   | || ||| |
Db        402 CCCCCCCCCCGCCCCCGCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCTCCGCCCCCCC 461

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
               ||| ||||| | ||   |||  | |  |||  |||  | ||||| |   |  ||| | |
Db        462 CCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCTCCCGCCACCCCCTCGCGCGCC 521

Qy        121 CCAGC 125
              |||| |
Db        522 CCACC 526


RESULT 6
US-10-767-701-8435
; Sequence 8435, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 8435
;   LENGTH: 678
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS39772_1
US-10-767-701-8435

  Query Match             44.8%;  Score 56;  DB 20;  Length 678;
  Best Local Similarity   68.8%;  Pred. No. 0.00017;
  Matches   77;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          2 CCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCG 61
              |||  |||| |||| ||||| |||| |||| |||| | ||| |      ||| | |
Db         29 CCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCGCCCCCGGCGCCCTGGGCCGGCCGGGG 88

Qy         62 CCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAG 113
              |||| |||| | |||  |||| | ||||||||||  | | |||||| | |
Db         89 CCCGGCCCCCCCCCGCCCCCGCCGCGGCCCGACCCCTTCCGGCCGCGGCCGG 140


RESULT 7
US-10-425-115-82176
; Sequence 82176, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 82176
;   LENGTH: 766
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_17495C.1
US-10-425-115-82176

  Query Match             44.6%;  Score 55.8;  DB 21;  Length 766;
  Best Local Similarity   70.1%;  Pred. No. 0.00018;
  Matches   75;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| |||| |||||||||||||||||| || || |||||  ||  | ||| |
Db        410 CCCCCCCCCCCCCCCCCGCCCCCACCCCCACCCCCCCCTCCCCCCCCCCCCCCCCCCAC 469

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
               ||| ||||| | |||  |||  | |  |||  ||| |  || || |
Db        470 CCCCACCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCC 516


RESULT 8
US-10-424-599-26988/c
; Sequence 26988, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 26988
;   LENGTH: 809
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(809)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_124370C.1
US-10-424-599-26988

  Query Match             44.5%;  Score 55.6;  DB 19;  Length 809;
  Best Local Similarity   63.2%;  Pred. No. 0.0002;
  Matches   79;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              || |  |  |  |||| ||||| ||||| ||||| ||||| |||||  ||  | ||| |
Db        662 CCGCCCCNGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNCCCCCCCCCCC 603

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
```

```
                 || | | | | ||   ||| | || ||||  ||    |||||||| ||   |  |||
Db        602 CNCCCCNCNCCCCCCCCCCCCCCCCCCGCCCCGCGCCCCCNCCGCCGCGCGCCCCCCCCGGC 543

Qy        121 CCAGC 125
              || ||
Db        542 CCCGC 538


RESULT 9
US-10-437-963-28390
; Sequence 28390, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 28390
;   LENGTH: 862
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(862)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_32995C.1
US-10-437-963-28390

  Query Match             44.3%;  Score 55.4;  DB 20;  Length 862;
  Best Local Similarity   69.2%;  Pred. No. 0.00021;
  Matches   74;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||| ||||  ||||  ||||| ||||| ||||| ||||| |||||  ||  | ||| |
Db        455 CCCCTCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 514

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
              ||| ||||| | ||   ||| ||| ||| |||   ||| || |
Db        515 CCCCCCCCCCCCTCCCCCCCCCCCTCCCCCCCCCCCCCNTCCCCCCC 561


RESULT 10
US-10-027-632-31508
; Sequence 31508, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31508
;   LENGTH: 956
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(956)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-31508

  Query Match             44.2%;  Score 55.2;  DB 14;  Length 956;
  Best Local Similarity   69.2%;  Pred. No. 0.00023;
  Matches   72;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          8 CCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCCCGCC 67
              | |  ||||  ||||| ||||| ||||| ||||| ||| | |||  | ||| | ||| ||
Db        428 CNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCGCCCCCCCCCCCCCCCCC 487

Qy         68 CCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGC 111
              ||||  || ||||| |   |||| |  |||   |||| |||| |
Db        488 CCCGGNCCCGGCCCCGGCNGGCCAGCGCCGCACCCGCGGCGGCC 531


RESULT 11
US-10-027-632-31508
; Sequence 31508, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31508
;   LENGTH: 956
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(956)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-31508

  Query Match             44.2%;  Score 55.2;  DB 18;  Length 956;
  Best Local Similarity   69.2%;  Pred. No. 0.00023;
  Matches   72;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          8 CCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCCCGCC 67
              | |  ||||  ||||| ||||| ||||| ||||| ||| | |||  | ||| | ||| ||
Db        428 CNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCGCCCCCCCCCCCCCCCCC 487
```

```
Qy         68 CCCGCGCCGGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGC 111
              ||||  || |||||  |   |||| |  |||   |||| |||| |
Db        488 CCCGGNCCCGGCCCCGGCNGGCCAGCGCCGCACCCGCGGCGGCC 531


RESULT 12
US-10-767-701-5654
; Sequence 5654, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 5654
;   LENGTH: 713
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2069_1
US-10-767-701-5654

  Query Match             44.0%;  Score 55;  DB 20;  Length 713;
  Best Local Similarity   66.4%;  Pred. No. 0.00028;
  Matches   79;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||  |||||  ||||| || || ||||| ||||| || || |||||  ||  | |||| |
Db         31 CCCTCGCCCCCCCCCCTTCCGCCCCCCCCGCCCCCCCCGCCCCCCCCTCCCCCCCCCCC 90

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGG 119
              |||| | | | | ||| |||| | |  |||   |||   |||||| |    ||||| |
Db         91 GCCCCCTCGCCCCCCGCGCCCCCCCCCCCCCCGCCGCCGCCGCGGGCCGCGGCGGCG 149


RESULT 13
US-10-425-115-93481
; Sequence 93481, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 93481
;   LENGTH: 581
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(581)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_185241C.1
US-10-425-115-93481

  Query Match             43.8%;  Score 54.8;  DB 21;  Length 581;
  Best Local Similarity   66.7%;  Pred. No. 0.00034;
  Matches   74;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              | ||  ||||  || |  | ||| ||||| ||||| ||||| ||| | ||| |
Db        400 CNCCCCCCCCCCCCNCNCCNCCCCCCCCCCCCCCCCCCCCCCCCCGGCCCCCCCCCCC 459

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGC 111
              ||| ||||| | ||   |||| | || |||  | |    || || |||||
Db        460 CCCCNCCCCCCCCCCCCCCCCGCCCCGCCCCCCCGCCCCCCCCCCCCGGGC 510


RESULT 14
US-10-425-115-63819
; Sequence 63819, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 63819
;   LENGTH: 788
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(788)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_158202C.1
US-10-425-115-63819

  Query Match             43.8%;  Score 54.8;  DB 21;  Length 788;
  Best Local Similarity   69.2%;  Pred. No. 0.0003;
  Matches   74;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||||  |||| |||||  ||| | ||||| ||||| ||||| |||||| ||| |
Db        677 CCCCCGCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCGGCCGGCCCCCCC 736

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGC 107
              |||| ||||| | ||   | |  | | || |  |||    || || |
Db        737 GCCCTCCCCCCCCCCCCCCGCCCCCCNGCGCTCCCCCCCCCCCCCCC 783


RESULT 15
US-10-767-701-9291
; Sequence 9291, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
;  APPLICANT: Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof For Plant Improvement
;  FILE REFERENCE: 38-21(53535)B
;  CURRENT APPLICATION NUMBER: US/10/767,701
;  CURRENT FILING DATE:  2004-01-29
;  NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 9291
;   LENGTH: 793
;   TYPE: DNA
;   ORGANISM: Sorghum bicolor
;   FEATURE:
;   OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS11175_1
US-10-767-701-9291

  Query Match             43.8%;  Score 54.8;  DB 20;  Length 793;
  Best Local Similarity   67.5%;  Pred. No. 0.0003;
  Matches   77;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;
```

```
Qy          9 CCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCCCGCCC 68
              ||  ||||  ||||| | ||| ||||| ||||| |||||  ||| ||||| | ||| |||
Db          9 CCCCCCCCCCCCCCCCCGCCCGCCCCCCCCCCCCCCCCCCCCCCGCCGCCCCCTCCCCCCC 68

Qy         69 CCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGCCC 122
              |||  ||  ||||  | || |||| ||| |   || | ||  | |  | ||| |
Db         69 CCGGCCCCAGCCCTCCCCGCCCCGCCCCAGCCGCGGCCCGCTCGGCGGCGGCGC 122
```

Search completed: November 8, 2005, 23:53:29
Job time : 162.744 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 22:16:08 ; Search time 645.819 Seconds
(without alignments)
7367.447 Million cell updates/sec

Title: US-09-463-542-1_COPY_1_125
Perfect score: 125
Sequence: 1 cccctgcccctgcccctgcc.....gcgggcaggcggggcccagc 125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	65	52.0	821	8	AQ893474	AQ893474 HS_4832_A
	2	64.4	51.5	1283	9	AG435406	AG435406 Mus muscu
c	3	63.4	50.7	982	6	CD329933	AGENCOURT
	4	63.2	50.6	425	2	BE551555	BE551555 hx97f12.x
	5	62.4	49.9	1866	3	CR622226	CR622226 full-leng
	6	62.2	49.8	996	5	BQ708087	AGENCOURT
	7	62	49.6	1024	6	CA975437	CA975437 AGENCOURT
	8	61.4	49.1	1674	9	CL078342	CL078342 CH216-149
c	9	61.2	49.0	843	9	CNS00CS1	AL059666 Drosophil
c	10	61	48.8	506	9	CNS027YY	AL185299 Tetraodon
	11	61	48.8	1003	9	CL475036	CL475036 SAIL_230_
c	12	61	48.8	1285	9	AG334095	AG334095 Mus muscu
	13	60.8	48.6	908	9	CNS006B4	AL064031 Drosophil
	14	60.8	48.6	1134	5	BQ651163	BQ651163 AGENCOURT
c	15	60.4	48.3	840	9	AG043467	AG043467 Pan trogl
	16	60.4	48.3	1101	9	CNS0153F	AL104949 Drosophil
	17	60.2	48.2	421	8	BH254057	BH254057 SALK_0159
c	18	60.2	48.2	566	9	CNS03JN0	AL247077 Tetraodon
	19	60.2	48.2	908	9	CL500140	CL500140 SAIL_679_
c	20	60.2	48.2	953	9	AG391885	AG391885 Mus muscu
	21	60.2	48.2	1014	9	CL492395	CL492395 SAIL_566_
c	22	60.2	48.2	1046	5	BQ952554	BQ952554 AGENCOURT
	23	60.2	48.2	1065	5	BU148615	BU148615 AGENCOURT
	24	60.2	48.2	1142	9	CL468510	CL468510 SAIL_1287
	25	60.2	48.2	1239	9	AG435545	AG435545 Mus muscu
	26	60.2	48.2	1406	9	AG435446	AG435446 Mus muscu
	27	60	48.0	921	9	CL465012	CL465012 SAIL_1232
	28	59.8	47.8	902	9	CL476702	CL476702 SAIL_258_
	29	59.8	47.8	1183	4	BM464236	BM464236 AGENCOURT
	30	59.6	47.7	899	9	CL495967	CL495967 SAIL_618_
c	31	59.6	47.7	925	9	CNS00CAZ	AL058951 Drosophil
	32	59.6	47.7	1157	4	BM466479	BM466479 AGENCOURT
	33	59.6	47.7	1313	4	BM552701	BM552701 AGENCOURT
	34	59.6	47.7	1372	4	BM461243	BM461243 AGENCOURT
c	35	59.6	47.7	1448	8	CC220110	CC220110 CH261-92F
c	36	59.4	47.5	244	9	AG081394	AG081394 Pan trogl
	37	59.4	47.5	715	5	BQ925271	BQ925271 AGENCOURT
c	38	59.4	47.5	870	9	AG407383	AG407383 Mus muscu
	39	59.4	47.5	899	9	CL493183	CL493183 SAIL_575_
c	40	59.4	47.5	1056	9	AG367644	AG367644 Mus muscu
c	41	59.4	47.5	1123	9	CL470578	CL470578 SAIL_144_
	42	59.4	47.5	1380	4	BM806217	BM806217 AGENCOURT
c	43	59.2	47.4	468	5	BQ608819	BQ608819 BRY_4732
c	44	59.2	47.4	613	9	CNS01HTD	AL144866 Anopheles
	45	59.2	47.4	823	2	BF866526	BF866526 963070A02

ALIGNMENTS

RESULT 1
AQ893474/c
LOCUS AQ893474 821 bp DNA linear GSS 10-NOV-1999
DEFINITION HS_4832_A2_C06_T7A CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=4832 Col=12 Row=E, genomic survey sequence.
ACCESSION AQ893474
VERSION AQ893474.1 GI:6349664
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 821)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 4832 row: E column: 12
Seq primer: T7
Class: BAC ends
High quality sequence stop: 821.
FEATURES Location/Qualifiers
source 1. .821
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=4832 Col=12 Row=E"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
ORIGIN

```
  Query Match            52.0%;  Score 65;  DB 8;  Length 821;
  Best Local Similarity  76.2%;  Pred. No. 0.056;
  Matches   80;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCCAGCCGGCGCCCGC 60
              ||||| ||||  |||| |||||| ||||| ||||| ||||| ||||| ||||||||||| |
Db        741 CCCCTCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCGCCGGCGCCCCC 682

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCC 105
              ||| || || | ||   |||||| | ||||  ||||   |||||
Db        681 CCCCCCCGCCCCCCCCCCCCCGGCCCCGCCCCCCCCGCCCCCGCC 637


RESULT 2
AG435406
LOCUS       AG435406               1283 bp    DNA     linear   GSS 03-JUN-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-313H22.T7, genomic survey
            sequence.
ACCESSION   AG435406
VERSION     AG435406.1  GI:48078469
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     BAC end Sequences of Library MSMg01
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1283)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1283
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-313H22.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match            51.5%;  Score 64.4;  DB 9;  Length 1283;
  Best Local Similarity  70.5%;  Pred. No. 0.065;
  Matches   86;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  |||| ||||  |||||| ||||| ||||| ||| | |||||   ||  | |||||
Db        908 CCCCCCCCCCCGCCCGCGCCCCCCCCCCCCCCCCCCCCCGCCCCCCCGCCCCCCCCCGC 967

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||||||||||| |||  |||| | |  |||| |||||  || || | | |  || | |
Db        968 CCCCGCCCCCGCCCCGCCCCCGCCCCCCCCCGCCCCGGCCCCCCCCCCGCCCCGCCGCCC 1027

Qy        121 CC 122
              ||
Db       1028 GC 1029


RESULT 3
CD329933/c
LOCUS       CD329933               982 bp    mRNA    linear   EST 28-MAY-2003
DEFINITION  AGENCOURT_14164463 NICHD_XGC_Brn1 Xenopus laevis cDNA clone
            IMAGE:6951625 5', mRNA sequence.
ACCESSION   CD329933
VERSION     CD329933.1  GI:31094264
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Xenopus.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement:
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14570  row: p  column: 24
            High quality sequence stop: 297.
FEATURES             Location/Qualifiers
     source          1. .982
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:6951625"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_Brn1"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5 kb. Constructed by Life
                     Technologies. Note: This is a Xenopus Gene Collection
                     (XGC) library."
ORIGIN

  Query Match            50.7%;  Score 63.4;  DB 6;  Length 982;
  Best Local Similarity  70.2%;  Pred. No. 0.1;
  Matches   85;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          2 CCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCG 61
              |||  |||  ||||| |||||  || || ||||| ||||  |||||  ||||| ||| |
Db        560 CCCCCCCCGCGCCCCCCCCCCCCCCCGCCCCCCCCCCCCCTCCCCCCCCCCGGCTCCCCCC 501

Qy         62 CCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGCC 121
              ||  |||| | ||||| || || ||||||   ||||  ||||||| |||  |||   ||
Db        500 GCCCGCCCCCCCCCGGGGCCCGCGCGGCCCCCGCCGGGCCCGCCGCCGGCGCGCGCCCCC 441

Qy        122 C 122
              |
Db        440 C 440


RESULT 4
BE551555
LOCUS       BE551555               425 bp    mRNA    linear   EST 10-AUG-2000
```

```
DEFINITION  hx97f12.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3195791 3'
            similar to contains element MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   BE551555
VERSION     BE551555.1  GI:9793338
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 206.
FEATURES             Location/Qualifiers
     source          1. .425
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3195791"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GC6"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
ORIGIN

  Query Match            50.6%;  Score 63.2;  DB 2;  Length 425;
  Best Local Similarity  68.8%;  Pred. No. 0.13;
  Matches   86;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| ||||| |||||||||||| ||||| ||||| |||||  ||  | ||| |
Db        269 CCCCCCCCCCCCCCCCTCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCC 328

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||| ||||| | |||||||||    |||| | |  ||  ||  |  |||   | ||| |
Db        329 CCCCCCCCCCTCCCCGGGCCCCTGCCGGCGGGCCGGGGGCCCCGCCGGGGGGCGAGGGCC 388

Qy        121 CCAGC 125
              || ||
Db        389 CCCGC 393


RESULT 5
CR622226
LOCUS       CR622226                1866 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DC004YO08 of Neuroblastoma Cot
            25-normalized of Homo sapiens (human).
ACCESSION   CR622226
VERSION     CR622226.1  GI:50503033
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1866)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1866)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .1866
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC004YO08"
                     /tissue_type="Neuroblastoma Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match            49.9%;  Score 62.4;  DB 3;  Length 1866;
  Best Local Similarity  98.4%;  Pred. No. 0.13;
  Matches   63;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         62 CCCGCCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGCTCCGCCGCGGGCAGGCGGGGCC 121
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db          1 CCCGCCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGCTCCGCCGCGGGCAGGCGGGGCC 60

Qy        122 CAGC 125
              ||||
Db         61 CAGC 64


RESULT 6
BQ708087
LOCUS       BQ708087                996 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8475084 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301404
            5', mRNA sequence.
ACCESSION   BQ708087
VERSION     BQ708087.1  GI:21846986
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 996)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Mark Watson
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2517  row: d  column: 13
            High quality sequence stop: 276.
FEATURES             Location/Qualifiers
```

```
     source          1. .996
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6301404"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_113"
                     /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             49.8%;  Score 62.2;  DB 5;  Length 996;
  Best Local Similarity   67.2%;  Pred. No. 0.16;
  Matches   82;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  | ||  ||||| | ||| ||||| ||||| ||||| ||     ||| |
Db        626 CCCCCCCCCCCCCNCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCNCCCCCCGCNNCNNCCCCC 685

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||||| ||||| | ||| |||| || |  | | | |   ||||||  || |||||  |
Db        686 CCCCGCCGCCGCGGCCGGCGCGGCGCGCCGGGCCGCCGCCGCGCCGCCCGCCGGCGGCCC 745

Qy        121 CC 122
              ||
Db        746 GC 747


RESULT 7
CA975437
LOCUS       CA975437               1024 bp    mRNA    linear   EST 06-JAN-2003
DEFINITION  AGENCOURT_11067725 NCI_CGAP_ZEmb3 Danio rerio cDNA clone
            IMAGE:6803474 5', mRNA sequence.
ACCESSION   CA975437
VERSION     CA975437.1  GI:27508091
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 1024)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Chi-Bin Chien
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14326  row: h  column: 01
            High quality sequence start: 40
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .1024
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:6803474"
                     /tissue_type="embryo"
                     /lab_host="DH10B (T1-resistant)"
                     /clone_lib="NCI_CGAP_ZEmb3"
                     /note="Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.1 kb. Constructed by J. Wang (Research Genetics,
                     Invitrogen Corp) from tissue donated by L. Zon (Harvard
                     University). Note: this is a NCI_CGAP Library."
ORIGIN

  Query Match             49.6%;  Score 62;  DB 6;  Length 1024;
  Best Local Similarity   68.6%;  Pred. No. 0.17;
  Matches   83;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  ||||  ||||| ||||| ||||| ||||| ||||   ||  | ||| |
Db        598 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCNCCCCCCCCCCCC 657

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||| ||||  | ||  |||| || | ||||| |||||  | | | ||||| |   | ||
Db        658 CCCCCCCCCNCCCCCCCGCCCCGCCCCGCCCCGCCCCGGCCCGGGCCCGGGCGGCGGCGGG 717

Qy        121 C 121
              |
Db        718 C 718


RESULT 8
CL078342
LOCUS       CL078342               1674 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-149D22_Sp6.1 CH216 Xenopus tropicalis genomic clone
            CH216-149D22, genomic survey sequence.
ACCESSION   CL078342
VERSION     CL078342.1  GI:40534255
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1674)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 300
            High quality sequence stop: 387.
FEATURES             Location/Qualifiers
     source          1. .1674
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-149D22"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             49.1%;  Score 61.4;  DB 9;  Length 1674;
  Best Local Similarity   72.1%;  Pred. No. 0.19;
  Matches   80;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  |||||  ||||  ||||| ||||| ||||| ||||| |||||  ||  |||||||
Db        190 CCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCGC 249

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGC 111
```

```
                 |||| ||||| | || | |||  | |  |||  |||    || ||||||
Db        250 GCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCGGCC 300


RESULT 9
CNS00CS1/c
LOCUS       CNS00CS1                 843 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR26H19 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL059666
VERSION     AL059666.1  GI:4947129
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR26H19"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             49.0%;  Score 61.2;  DB 9;  Length 843;
  Best Local Similarity   60.5%;  Pred. No. 0.24;
  Matches   75;  Conservative   11;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  |:||  ||||  ||||  ||||| ||||| ||||| |||||  ||  | ||| |
Db        274 CCCCSCCSCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 215

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
               ||| ||||| | ||   ||| |: || ||:: ::|    || ::::|||  |: |||||
Db        214 CCCCCCCCCCCCCCCCCCCCCCGSNCGCCCSSSSSCCNGNCCCSSSSGGGGGGSGGGGGC 155

Qy        121 CCAG 124
               | |
Db        154 GCSG 151


RESULT 10
CNS027YY/c
LOCUS       CNS027YY                 506 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            243J12 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL185299
VERSION     AL185299.1  GI:7823403
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 506)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="243J12"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG243DE06SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             48.8%;  Score 61;  DB 9;  Length 506;
  Best Local Similarity   52.1%;  Pred. No. 0.29;
  Matches   62;  Conservative   30;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  ||||  ||||| |:||| |:||: ||||| |:|||  ||: ::|||:|
Db        344 CCCCCCCCCCCCCCCCCCCCCCCCCMCCCCCMCCSCCCCCCCCSCCCCCCCSCSSCCCSC 285

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGG 119
              |||:||::| ::|| ::::|:|:  | :| : :| ||   :|::::|:|  ||:|||:
Db        284 CCCCSCCSSCCSSCCCSSSSCBGSSGGCSCGSGBCGGGCGGSGBVBSGSGGCGGVGGGB 226


RESULT 11
CL475036
LOCUS       CL475036                 1003 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_230_F07.v1 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_230_F07.v1, genomic survey sequence.
ACCESSION   CL475036
VERSION     CL475036.1  GI:45940747
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
```

```
REFERENCE   1  (bases 1 to 1003)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
  MEDLINE   22356987
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS810670; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1003
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_230_F07.v1"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match             48.8%;  Score 61;  DB 9;  Length 1003;
  Best Local Similarity   67.2%;  Pred. No. 0.25;
  Matches   82;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||  |  |||| |||||| ||||| ||||| ||| |  | | | || ||
Db        876 CCCCCCCCGCCCCCCCCGCCCCCGCCCCCCCCCCCCCCCCCCNCCCGCCCCNGCCCCCNGC 935

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||||  ||||| ||| ||||| | |  | || ||||   ||||  || ||   |  ||
Db        936 CCCCGCNCCCGCCCCGCGCCCGCCGCCCCGCGCCCCGCCCCCGCGCCGCGCCCACNCCGC 995

Qy        121 CC 122
              ||
Db        996 GC 997


RESULT 12
AG334095/c
LOCUS       AG334095               1285 bp    DNA     linear   GSS 02-JUN-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-124N12.TJ, genomic survey
            sequence.
ACCESSION   AG334095
VERSION     AG334095.1  GI:47907405
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     BAC end Sequences of Library MSMg01
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1285)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1285
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-124N12.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             48.8%;  Score 61;  DB 9;  Length 1285;
  Best Local Similarity   67.2%;  Pred. No. 0.24;
  Matches   82;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||||  ||||  ||||| ||||| ||||| ||||| |||   ||||||
Db        788 CCCCCGCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCGGCCGC 729

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||| |||||||  |   |||  | |  |||  |||    || |  || |   |||| |
Db        728 CCCCCCCCCCGCCGCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCNNCGCGNCCGCGGCNC 669

Qy        121 CC 122
              ||
Db        668 CC 667


RESULT 13
CNS006B4
LOCUS       CNS006B4                908 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR13L17 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064031
VERSION     AL064031.1  GI:4944211
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 908)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
```

```
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .908
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR13L17"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             48.6%;  Score 60.8;  DB 9;  Length 908;
  Best Local Similarity   66.4%;  Pred. No. 0.28;
  Matches   83;  Conservative    2;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||||  ||||  | ||| ||||: ||||| ||||| |||||  || || |||||
Db        270 CCCCCGCCCCGCCCCCCCCCGCCCNCCCCSCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCGC 329

Qy         61 GCCCGCCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||: ||||| | ||    ||| | | ||||| |||     | ||||   |  || | ||
Db        330 CCCSCCCCCCCCCCCCCCGCCGCCCCCGCCCGCCCCCCCCGCCCCGCCCCCCCCGCCGCGC 389

Qy        121 CCAGC 125
              || ||
Db        390 CCCGC 394


RESULT 14
BQ651163
LOCUS       BQ651163                1134 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8488514 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6296195
            5', mRNA sequence.
ACCESSION   BQ651163
VERSION     BQ651163.1  GI:21775335
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1134)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2503  row: k  column: 12
            High quality sequence start: 17
            High quality sequence stop: 135.
FEATURES             Location/Qualifiers
     source          1. .1134
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6296195"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match             48.6%;  Score 60.8;  DB 5;  Length 1134;
  Best Local Similarity   66.1%;  Pred. No. 0.27;
  Matches   80;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| |||||   |||| |||||   ||| ||||| ||| |  ||  | | | |
Db        330 CCCCCCCCCCGGCCCCGNNCCCCNCCCCCNNNCCCCCCCCCCCCCNCCCCCCCCCCNCCC 389

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||| ||||| ||||| | ||| | || ||    ||    |||||||| || ||||   |
Db        390 CCCCCCCCCCCCGCCGCGGCCGCCGCGCCCGCGGCCCCCCCCGCCGCGCGCNGGCGCCCC 449

Qy        121 C 121
              |
Db        450 C 450


RESULT 15
AG043467/c
LOCUS       AG043467                840 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-021M16.F, genomic survey sequence.
ACCESSION   AG043467
VERSION     AG043467.1  GI:16572192
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 840)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-021M16.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             48.3%;  Score 60.4;  DB 9;  Length 840;
  Best Local Similarity   66.1%;  Pred. No. 0.33;
  Matches   82;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
```

```
            ||||   |||  || |  ||||| ||||| ||||| ||||| || ||  ||  | ||| |
Db        544 CCCCCCGCCCCCCCNCCCCCCCCCCCCCCCNCCCCCNCCCCCCCCCGCCCCCCNNCCCCCCC 485

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||||||  |   | ||| |||||  | || |||  |||     |  ||||| || |||| ||
Db        484 CCCCGCCGCGCCCCCGCGCCCGCCCCGCCCCCNCCCCCCGCGCCCGCGCGCCGGCGCGGG 425

Qy        121 CCAG 124
                | |
Db        424 GCCG 421
```

Search completed: November 9, 2005, 04:17:46
Job time : 649.819 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:52:53 ; Search time 4078.01 Seconds
(without alignments)
9232.381 Million cell updates/sec

Title: US-09-463-542-34_COPY_368_1144
Perfect score: 777
Sequence: 1 tcatgtaggtaagactgtgt.....gtctctttttattgttaag 777

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	765.8	98.6	135675	9	AC093174	AC093174 Homo sapi
	2	765.8	98.6	148828	9	AY157024	AY157024 Homo sapi
	3	765.8	98.6	180963	9	AC091492	AC091492 Homo sapi
c	4	765.8	98.6	221402	2	AC015995	AC015995 Homo sapi
	5	764.2	98.4	851	9	AF548352	AF548352 Homo sapi
c	6	565.4	72.8	186028	2	AC151849	AC151849 Callithri
	7	421	54.2	65160	2	AC135179	AC135179 Homo sapi
	8	130.6	16.8	212216	2	AC131898	AC131898 Oryctolag
	9	103.6	13.3	185737	2	AC125447	AC125447 Mus muscu
c	10	88.6	11.4	191540	2	AC136055	AC136055 Rattus no
c	11	88.6	11.4	245724	2	AC091418	AC091418 Rattus no
c	12	88.6	11.4	250169	2	AC120668	AC120668 Rattus no
	13	68.6	8.8	154195	2	AC142242	AC142242 Atelerix
	14	53	6.8	183103	5	BX465190	BX465190 Zebrafish
	15	52.8	6.8	145160	2	CR847935	CR847935 Danio rer
	16	51.8	6.7	166021	2	AC117751	AC117751 Mus muscu
c	17	50.8	6.5	153477	2	AC006278	AC006278 Plasmodiu
c	18	50.8	6.5	251551	3	AE014844	AE014844 Plasmodiu
	19	50	6.4	5771	6	AX344664	AX344664 Sequence
	20	50	6.4	125020	9	AF429315	AF429315 Homo sapi
c	21	49.6	6.4	95996	8	ATF18P9	AL138654 Arabidops
	22	49.4	6.4	175188	2	BX897715	BX897715 Danio rer
	23	49.2	6.3	18683	6	AX281291	AX281291 Sequence
	24	49.2	6.3	18683	6	AX345214	AX345214 Sequence
c	25	49.2	6.3	156276	2	CR769765	CR769765 Danio rer
c	26	49.2	6.3	175234	5	BX547997	BX547997 Zebrafish
c	27	49.2	6.3	241364	2	BX571898	BX571898 Danio rer
	28	49	6.3	11422	6	AX345121	AX345121 Sequence
	29	49	6.3	11422	6	AX348323	AX348323 Sequence
	30	49	6.3	138121	9	AC114806	AC114806 Homo sapi
	31	48.8	6.3	110000	2	PFMAL8P1_08	Continuation (9 of
	32	48.8	6.3	163443	2	AC006280	AC006280 Plasmodiu
	33	48.8	6.3	196149	2	AC004709	AC004709 Plasmodiu
c	34	48.8	6.3	252650	3	AE014847	AE014847 Plasmodiu
	35	48.6	6.3	179824	9	AC096708	AC096708 Homo sapi
	36	48.4	6.2	13123	6	AX281322	AX281322 Sequence
	37	48.4	6.2	13123	6	AX344419	AX344419 Sequence
	38	48	6.2	92725	9	AC010350	AC010350 Homo sapi
	39	48	6.2	152706	9	AC008568	AC008568 Homo sapi
c	40	48	6.2	152707	9	AC008876	AC008876 Homo sapi
	41	48	6.2	349980	6	AX344565	AX344565 Sequence
	42	47.6	6.1	958	3	NAM556112	AJ556112 Necator a
	43	47.6	6.1	958	3	NAM556113	AJ556113 Necator a
	44	47.6	6.1	958	3	NAM556114	AJ556114 Necator a
	45	47.6	6.1	958	3	NAM556116	AJ556116 Necator a

ALIGNMENTS

RESULT 1
AC093174
LOCUS AC093174 135675 bp DNA linear PRI 08-NOV-2002
DEFINITION Homo sapiens chromosome 3 clone RP11-167M22 map 3p, complete sequence.
ACCESSION AC093174
VERSION AC093174.2 GI:24796717
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 135675)
AUTHORS Wu,Q., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L., Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y., Wu,D., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and Yang,H.
TITLE Chromosome 3p genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 135675)
AUTHORS Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L., Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J., Wang,J., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and Yang,H.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-2001) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China

```
REFERENCE   3  (bases 1 to 135675)
  AUTHORS   Wu,Q., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J.,
            Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D.,
            Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C.,
            Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B.,
            Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y.,
            Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X.,
            Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and
            Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
COMMENT     On Nov 8, 2002 this sequence version replaced gi:15148930.
            --------------Genome Center
            Center:Beijing Center
            Center code:Beijing
            Website:http://hgc.igtp.ac.cn
            http://www.genomics.org.cn
            Contact:hgc@igtp.ac.cn
            -------------- Project Information
            Center project name:1% project
            Center clone name: RP11-167M22
            -------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator: ET 55% of reads
            Chemistry: Dye-terminator Big Dye; 45% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 0 bases at least Q40
            Consensus quality: 0 bases at least Q30
            Consensus quality: 6 bases at least Q20
            Insert size: 3392; sum-of-contigs
            Quality coverage:   0.00x in Q20 bases;sum-of-contigs
            --------------.
FEATURES             Location/Qualifiers
     source          1. .135675
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3p"
                     /clone="RP11-167M22"
ORIGIN

  Query Match             98.6%;  Score 765.8;  DB 9;  Length 135675;
  Best Local Similarity   99.1%;  Pred. No. 6.5e-160;
  Matches  770;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41828 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 41887

Qy         61 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41888 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 41947

Qy        121 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41948 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 42007

Qy        181 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42008 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 42067

Qy        241 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42068 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 42127

Qy        301 GTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTC 360
              |||||||||||||||||||||||||  ||||||||||||||||||||| || ||

Db      42128 GTGTCATCTAGAATGAAGTTATATTTTACTAAGGATCGTTTTTGCCATGTATAACTTCTC 42187

Qy        361 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42188 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 42247

Qy        421 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42248 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 42307

Qy        481 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAAT 540
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      42308 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTCTAATAGATGTTTCTTTTATGAAATAAT 42367

Qy        541 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      42368 TTTTAAAAGATTCTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 42427

Qy        601 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42428 ACTTCTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 42487

Qy        661 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42488 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 42547

Qy        721 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42548 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 42604


RESULT 2
AY157024
LOCUS       AY157024               148828 bp    DNA     linear   PRI 16-JUN-2004
DEFINITION  Homo sapiens peroxisome proliferative activated receptor gamma
            (PPARG) gene, complete cds.
ACCESSION   AY157024
VERSION     AY157024.2  GI:48762804
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148828)
  AUTHORS   Rieder,M.J., da Ponte,S.H., Kuldanek,S.A., Rajkumar,N., Smith,J.D.,
            Toth,E.J. and Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2002) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
REFERENCE   2  (bases 1 to 148828)
  AUTHORS   Rieder,M.J., Daniels,R.L., da Ponte,S.H., Hastings,N.C.,
            Ahearn,M.O., Rajkumar,N., Yi,Q. and Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-2004) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
  REMARK    Sequence update by submitter
COMMENT     On Jun 16, 2004 this sequence version replaced gi:23953882.
            To cite this work please use:  SeattleSNPs. NHLBI Program for
            Genomic Applications, UW-FHCRC, Seattle, WA
            (URL:http://pga.mbt.washington.edu).
FEATURES             Location/Qualifiers
     source          1. .148828
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     repeat_region   9. .59
                     /rpt_family="L1"
                     /rpt_type=dispersed
     variation       54
                     /frequency="0.12"
                     /replace="g"
     repeat_region   60. .351
                     /rpt_family="Alu"
```

```
                /rpt_type=dispersed
variation       147
                /frequency="0.03"
                /replace="t"
variation       252
                /frequency="0.03"
                /replace="g"
repeat_region   352. .431
                /rpt_family="L1"
                /rpt_type=dispersed
variation       374
                /frequency="0.05"
                /replace="c"
repeat_region   435. .759
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1555
                /frequency="0.02"
                /replace="t"
misc_feature    1754. .1925
                /note="Region not scanned for variation"
repeat_region   1784. .2127
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   2032. .2632
                /rpt_family="L1"
                /rpt_type=dispersed
gene            2048. .148443
                /gene="PPARG"
mRNA            join(2048. .2143,26500. .26573,93809. .94036,95427. .95596,
                106715. .106853,119980. .120179,130803. .131253,
                147997. .148443)
                /gene="PPARG"
                /product="peroxisome proliferative activated receptor
                gamma"
variation       2076
                /gene="PPARG"
                /frequency="0.15"
                /replace="t"
variation       2107
                /gene="PPARG"
                /frequency="0.01"
                /replace="a"
variation       2186
                /gene="PPARG"
                /frequency="0.22"
                /replace="c"
variation       2382
                /gene="PPARG"
                /frequency="0.01"
                /replace="t"
variation       2407
                /gene="PPARG"
                /frequency="0.05"
                /replace="t"
variation       2499
                /gene="PPARG"
                /frequency="0.04"
                /replace="t"
variation       2767
                /gene="PPARG"
                /frequency="0.07"
                /replace="g"
variation       3035
                /gene="PPARG"
                /frequency="0.01"
                /replace="a"
variation       3198
                /gene="PPARG"
                /frequency="0.04"
                /replace="t"
variation       3354
                /gene="PPARG"
                /frequency="0.01"
                /replace=""
variation       4160
                /gene="PPARG"
                /frequency="0.01"
                /replace="c"
misc_feature    4220. .25559
                /gene="PPARG"
                /note="Region not scanned for variation"
repeat_region   5952. .6004
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   8735. .9019
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   9637. .9935
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   10092. .10404
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   10423. .10537
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   11193. .11356
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   12140. .12449
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   12949. .13255
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   13277. .13383
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   16397. .16653
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   17063. .17196
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   17325. .17622
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   17653. .17733
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   17766. .17984
                /rpt_family="CR1"
                /rpt_type=dispersed
repeat_region   18787. .18845
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   19727. .19804
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   19757. .19813
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   20083. .20181
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   20182. .20524
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   20525. .20660
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   20937. .21239
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   21961. .22204
```

```
                     /rpt_family="L2"
                     /rpt_type=dispersed
     repeat_region   22205. .22492
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   22493. .22581
                     /rpt_family="L2"
                     /rpt_type=dispersed
     repeat_region   22582. .22916
                     /rpt_family="MER1_type"
                     /rpt_type=dispersed
     repeat_region   22917. .23104
                     /rpt_family="L2"
                     /rpt_type=dispersed

  Query Match             98.6%;  Score 765.8;  DB 9;  Length 148828;
  Best Local Similarity   99.1%;  Pred. No. 6.5e-160;
  Matches  770;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25723 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 25782

Qy         61 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25783 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 25842

Qy        121 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25843 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 25902

Qy        181 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25903 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 25962

Qy        241 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25963 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 26022

Qy        301 GTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTC 360
              |||||||||||||||||||||||||  ||||||||||||||||||||||| || ||
Db      26023 GTGTCATCTAGAATGAAGTTATATTTTACTAAGGATCGTTTTTGCCATGTATAACTTCTC 26082

Qy        361 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26083 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 26142

Qy        421 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26143 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 26202

Qy        481 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAAT 540
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      26203 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTCTAATAGATGTTTCTTTTATGAAATAAT 26262

Qy        541 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      26263 TTTTAAAAGATTCTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 26322

Qy        601 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26323 ACTTCTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 26382

Qy        661 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26383 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 26442

Qy        721 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26443 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 26499


RESULT 3
```

```
AC091492
LOCUS       AC091492              180963 bp    DNA     linear   PRI 20-DEC-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-335I9 map 3p, complete
            sequence.
ACCESSION   AC091492
VERSION     AC091492.2  GI:24796732
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180963)
  AUTHORS   Bao,W., Bao,J., Bao,Q., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
            Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X.,
            Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Chromosome 3p genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180963)
  AUTHORS   Bao,W., Bao,J., Bao,Q., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2001) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
REFERENCE   3  (bases 1 to 180963)
  AUTHORS   Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
REFERENCE   4  (bases 1 to 180963)
  AUTHORS   Bao,W., Bao,J., Bao,Q., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-2002) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
```

```
COMMENT     On Nov 8, 2002 this sequence version replaced gi:13794231.
            -------------Genome Center
            Center:Beijing Center
            Center code:Beijing
            Website:http://hgc.igtp.ac.cn
            http://www.genomics.org.cn
            Contact:hgc@igtp.ac.cn
            -------------- Project Information
            Center project name:1% project
            Center clone name: RP11-335I9
            -------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator: ET 55% of reads
            Chemistry: Dye-terminator Big Dye; 45% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 369 bases at least Q40
            Consensus quality: 571 bases at least Q30
            Consensus quality: 755 bases at least Q20
            Insert size: 773; sum-of-contigs
            Quality coverage:   1.48x in Q20 bases;sum-of-contigs
            --------------.
FEATURES             Location/Qualifiers
     source          1. .180963
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3p"
                     /clone="RP11-335I9"
ORIGIN

  Query Match             98.6%;  Score 765.8;  DB 9;  Length 180963;
  Best Local Similarity   99.1%;  Pred. No. 6.3e-160;
  Matches  770;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161848 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 161907

Qy         61 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161908 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 161967

Qy        121 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161968 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 162027

Qy        181 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     162028 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 162087

Qy        241 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     162088 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 162147

Qy        301 GTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTC 360
              |||||||||||||||||||||||||  |||||||||||||||||||||| || ||
Db     162148 GTGTCATCTAGAATGAAGTTATATTTTACTAAGGATCGTTTTTGCCATGTATAACTTCTC 162207

Qy        361 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     162208 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 162267

Qy        421 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     162268 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 162327

Qy        481 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAAT 540
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     162328 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTCTAATAGATGTTTCTTTTATGAAATAAT 162387

Qy        541 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||

Db     162388 TTTTAAAAGATTCTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 162447

Qy        601 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     162448 ACTTCTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 162507

Qy        661 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     162508 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 162567

Qy        721 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     162568 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 162624


RESULT 4
AC015995/c
LOCUS       AC015995               221402 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-335I9 map 3, WORKING DRAFT
            SEQUENCE, 36 unordered pieces.
ACCESSION   AC015995
VERSION     AC015995.5  GI:9966961
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 221402)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-335I9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 221402)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 221402)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
```

```
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 5, 2000 this sequence version replaced gi:7329342.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2392
                Center clone name: 335_I_9
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 97% of reads
                Sequencing vector: Plasmid; n/a; %-0.f%% of reads
                2.51630941286114Chemistry: Dye-terminator Big Dye; 100% of
            reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 188096 bases at least Q40
                Consensus quality: 198983 bases at least Q30
                Consensus quality: 206638 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 217902; sum-of-contigs
                Quality coverage: 6.0 in Q20 bases; agarose-fp
                Quality cov.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 36 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3984: contig of 3984 bp in length
            *     3985     4084: gap of 100 bp
            *     4085     5217: contig of 1133 bp in length
            *     5218     5317: gap of 100 bp
            *     5318     6422: contig of 1105 bp in length
            *     6423     6522: gap of 100 bp
            *     6523     7602: contig of 1080 bp in length
            *     7603     7702: gap of 100 bp
            *     7703     8882: contig of 1180 bp in length
            *     8883     8982: gap of 100 bp
            *     8983    10209: contig of 1227 bp in length
            *    10210    10309: gap of 100 bp
            *    10310    12072: contig of 1763 bp in length
            *    12073    12172: gap of 100 bp
            *    12173    13428: contig of 1256 bp in length
            *    13429    13528: gap of 100 bp
            *    13529    14796: contig of 1268 bp in length
            *    14797    14896: gap of 100 bp
            *    14897    16049: contig of 1153 bp in length
            *    16050    16149: gap of 100 bp
            *    16150    17583: contig of 1434 bp in length
            *    17584    17683: gap of 100 bp
            *    17684    18808: contig of 1125 bp in length
            *    18809    18908: gap of 100 bp
            *    18909    20356: contig of 1448 bp in length
            *    20357    20456: gap of 100 bp
            *    20457    22555: contig of 2099 bp in length
            *    22556    22655: gap of 100 bp
            *    22656    24625: contig of 1970 bp in length
            *    24626    24725: gap of 100 bp
            *    24726    26352: contig of 1627 bp in length
            *    26353    26452: gap of 100 bp
            *    26453    28366: contig of 1914 bp in length
            *    28367    28466: gap of 100 bp
            *    28467    29704: contig of 1238 bp in length
            *    29705    29804: gap of 100 bp
            *    29805    31337: contig of 1533 bp in length
            *    31338    31437: gap of 100 bp
            *    31438    32941: contig of 1504 bp in length
            *    32942    33041: gap of 100 bp
            *    33042    34690: contig of 1649 bp in length
            *    34691    34790: gap of 100 bp
            *    34791    36510: contig of 1720 bp in length
            *    36511    36610: gap of 100 bp
            *    36611    39065: contig of 2455 bp in length
            *    39066    39165: gap of 100 bp
            *    39166    41551: contig of 2386 bp in length
            *    41552    41651: gap of 100 bp
            *    41652    44456: contig of 2805 bp in length
            *    44457    44556: gap of 100 bp
            *    44557    46379: contig of 1823 bp in length
            *    46380    46479: gap of 100 bp
            *    46480    49222: contig of 2743 bp in length
            *    49223    49322: gap of 100 bp
            *    49323    52311: contig of 2989 bp in length
            *    52312    52411: gap of 100 bp
            *    52412    57964: contig of 5553 bp in length
            *    57965    58064: gap of 100 bp
            *    58065    63519: contig of 5455 bp in length
            *    63520    63619: gap of 100 bp
            *    63620    70598: contig of 6979 bp in length
            *    70599    70698: gap of 100 bp
            *    70699    83449: contig of 12751 bp in length
            *    83450    83549: gap of 100 bp
            *    83550   100089: contig of 16540 bp in length
            *   100090   100189: gap of 100 bp
            *   100190   120532: contig of 20343 bp in length
            *   120533   120632: gap of 100 bp
            *   120633   160594: contig of 39962 bp in length
            *   160595   160694: gap of 100 bp
            *   160695   221402: contig of 60708 bp in length.
FEATURES             Location/Qualifiers
     source          1. .221402
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-335I9"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .3984
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4085. .5217
                     /note="assembly_fragment"
     misc_feature    5318. .6422
                     /note="assembly_fragment"
     misc_feature    6523. .7602
                     /note="assembly_fragment"
     misc_feature    7703. .8882
                     /note="assembly_fragment"
     misc_feature    8983. .10209
                     /note="assembly_fragment"
     misc_feature    10310. .12072
                     /note="assembly_fragment"
     misc_feature    12173. .13428
                     /note="assembly_fragment"
     misc_feature    13529. .14796
                     /note="assembly_fragment"
     misc_feature    14897. .16049
                     /note="assembly_fragment"
     misc_feature    16150. .17583
                     /note="assembly_fragment"
     misc_feature    17684. .18808
                     /note="assembly_fragment"
     misc_feature    18909. .20356
                     /note="assembly_fragment"
     misc_feature    20457. .22555
                     /note="assembly_fragment"

  Query Match             98.6%;  Score 765.8;  DB 2;  Length 221402;
```

```
  Best Local Similarity   99.1%;  Pred. No. 6.1e-160;
  Matches  770;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      98593 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 98534

Qy         61 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      98533 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 98474

Qy        121 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      98473 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 98414

Qy        181 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      98413 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 98354

Qy        241 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      98353 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 98294

Qy        301 GTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTC 360
              |||||||||||||||||||||||||  ||||||||||||||||||||| || ||
Db      98293 GTGTCATCTAGAATGAAGTTATATTTTACTAAGGATCGTTTTTGCCATGTATAACTTCTC 98234

Qy        361 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      98233 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 98174

Qy        421 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      98173 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 98114

Qy        481 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAAT 540
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      98113 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTCTAATAGATGTTTCTTTTATGAAATAAT 98054

Qy        541 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      98053 TTTTAAAAGATTCTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 97994

Qy        601 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97993 ACTTCTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 97934

Qy        661 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97933 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 97874

Qy        721 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97873 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 97817


RESULT 5
AF548352
LOCUS       AF548352                851 bp    DNA     linear   PRI 01-MAR-2003
DEFINITION  Homo sapiens peroxisome proliferator-activated receptor gamma 3
            (PPARG) gene, promoter region and exon A2.
ACCESSION   AF548352
VERSION     AF548352.1  GI:28626258
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 851)
  AUTHORS   Meirhaeghe,A., Fajas,L., Gouilleux,F., Cottel,D., Helbecque,N.,
            Auwerx,J. and Amouyel,P.
  TITLE     A functional polymorphism in a STAT5B site of the human PPAR gamma
            3 gene promoter affects height and lipid metabolism in a French
            population
  JOURNAL   Arterioscler. Thromb. Vasc. Biol. (2003) In press
REFERENCE   2  (bases 1 to 851)
  AUTHORS   Meirhaeghe,A., Fajas,L., Helbecque,N., Auwerx,J. and Amouyel,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2002) INSERM U508, Institut Pasteur de Lille, 1
            rue du Pr Calmette - BP 245, Lille 59019, France
FEATURES             Location/Qualifiers
     source          1. .851
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3p25"
     gene            <1. .>851
                     /gene="PPARG"
                     /note="PPARg"
     promoter        <1. .777
                     /gene="PPARG"
                     /note="promoter 3"
     variation       97
                     /gene="PPARG"
                     /note="SNP"
                     /replace="g"
     variation       225
                     /gene="PPARG"
                     /note="SNP"
                     /replace="c"
     mRNA            778. .>851
                     /gene="PPARG"
                     /product="peroxisome proliferator-activated receptor gamma
                     3"
     exon            778. .851
                     /gene="PPARG"
                     /number=A2
ORIGIN

  Query Match             98.4%;  Score 764.2;  DB 9;  Length 851;
  Best Local Similarity   99.0%;  Pred. No. 3.1e-159;
  Matches  769;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60

Qy         61 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 120

Qy        121 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 180

Qy        181 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 240

Qy        241 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 300

Qy        301 GTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTC 360
              |||||||||||||||||||||||||  ||||||||||||||||||||| || ||
Db        301 GTGTCATCTAGAATGAAGTTATATTTTACTAAGGATCGTTTTTGCCATGTATAACTTCTC 360

Qy        361 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 420

Qy        421 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 480
```

```
Qy        481 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAAT 540
              ||||||||||||||||||||||||||||| | ||||||||||||||||||||||||||||
Db        481 GTGATGGAAATAGGAAAGTAGGTGAAGTGAATCTAATAGATGTTTCTTTTATGAAATAAT 540

Qy        541 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTTTAAAAGATTCTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600

Qy        601 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACTTCTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660

Qy        661 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 720

Qy        721 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777
```

RESULT 6
AC151849/c
LOCUS AC151849 186028 bp DNA linear HTG 08-OCT-2004
DEFINITION Callithrix jacchus clone CH259-42F14, WORKING DRAFT SEQUENCE, 2 ordered pieces.
ACCESSION AC151849
VERSION AC151849.1 GI:53983836
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Callithrix jacchus (white-tufted-ear marmoset)
ORGANISM Callithrix jacchus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
REFERENCE 1 (bases 1 to 186028)
AUTHORS Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R., Peng,Z., Malinov,I. and Rubin,E.M.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 186028)
AUTHORS Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R., Peng,Z., Malinov,I. and Rubin,E.M.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-2004) Genome Sciences, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
COMMENT
Sequence Produced by Berkeley PGA
Web site: http://pga.lbl.gov
Center Code: PGABERK
Center Project Name: J108-42F14
Bac Clone Name: CH259-42F14

This sequence has been compared to sequences of other species using Vista (http://www-gsd.lbl.gov/VISTA). The results can be viewed at:
http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=PPARG

The order-orientation of the draft sequence was accomplished by using:
Avid (http://baboon.math.berkeley.edu/mavid) and paired end information.

Funding agent: Programs for Genomic Applications (NHLBI)

Summary Statistics:
Sequencing vector: Plasmid; pUC18
Chemistry: Dye-terminator Big Dye
Assembly program: Phrap version 0.990329.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 26679: contig of 26679 bp in length
* 26680 26779: gap of unknown length
* 26780 186028: contig of 159249 bp in length.
FEATURES Location/Qualifiers
source 1. .186028
/organism="Callithrix jacchus"
/mol_type="genomic DNA"
/db_xref="taxon:9483"
/clone="CH259-42F14"
ORIGIN

Query Match 72.8%; Score 565.4; DB 2; Length 186028;
Best Local Similarity 87.8%; Pred. No. 1.8e-115;
Matches 663; Conservative 0; Mismatches 86; Indels 6; Gaps 4;

```
Qy         18 TGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGATGATAAGGCTTTTGGCA 77
              || ||||||| | |||| ||||||||| ||||| | ||||||||||||||||||||||||
Db     123605 TGCAGAATGTTGAGTCTTGATGTTGGCACTATTTAGGCCCTGATGATAAGGCTTTTGGCA 123546

Qy         78 TTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGGATCCTTGAGCCTTTCAT 137
              |||||||| ||||||||| |||||||||||||||||||||||| ||| ||||||||||
Db     123545 CTAGATGCTATTTTGTCTTGATGGAAAATACAGCTATTCTAGGATTCTTAAGCCTTTCAT 123486

Qy        138 AAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACTTAGATGATCTGCTCTCT 197
              |||||||||||||||||||||||||||||||||| || || |||||||||    ||||| |||||
Db     123485 AAGAGATAAGGTTGTGAATCCTAAGACCCTGGGCCCTTTTACTTAGGCAATCTGTTCTCT 123426

Qy        198 GGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCATTTGCCTTGCCTAAGTG 257
              | || ||   ||||| |||||||||||| || | |   ||||||||||||||||||  ||
Db     123425 TGCTCATC---TGAAAGGTCTGCTTCGTAAGTGTTACACTGCATTTGCCTTGCCTACATG 123369

Qy        258 G-TGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCATGTGTCATCTAGAATGA 316
              | || |||||||| |||||||||||  ||||||| ||||||||||||||||||||||||
Db     123368 GTTGAGGCACACATCTGTACTGTCATTTTAGGCTAAATAACCATGTGTCATCTAGAATGA 123309

Qy        317 AGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAG 376
              |||||||||||  |||||| |||||||| ||||||| || ||| |||||||| |||
Db     123308 AGTTATATTTTATTAAGGATTGTTTTTGCCGTGTATAAGTTATCACACATTAACCTTCGA 123249

Qy        377 GGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGT 436
              |||||||||||||||||||||| |||||||||||||  ||||||| |||||||| |
Db     123248 GGTTATTAATCCTTTTAAGGTCTAATTTTTCTTAAGTCTGCACAGTAATGGAGGTATCAT 123189

Qy        437 CATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAA 496
              |||||||||||||||||||| |||||||||  |||||||||||| ||||||||||
Db     123188 CATTCATGTGACATAAAAGATGAAAAGGGGCTTAGTTCATGTTAGTGACGGAAATAGGAA 123129

Qy        497 AGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCC 556
              ||||||| |||||| | ||||| |||| |||||||||||||||||||||||| || |||
Db     123128 AGTAGGTAAAGTGAATCTAATACATGTCTCTTTTATGAAATAATTTTTAAAAGGTTCTCC 123069

Qy        557 AGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGA 616
              ||||||| |||||||||||||||||||||| |||||||||||| | ||||||| |
Db     123068 AGCCCTGTATGATTTATGATGAATCATTTTGTGGTTTGTTAGTTACTTCTCGAGAATATA 123009

Qy        617 AAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAAT 676
              |||||||||| |||||||||||||||| |||||||||||||||||||||| | ||||
Db     123008 AAGCATTGTA-GCTCAGGGAAAGCAAACGTTCAGAATGAAATCCAATAGAGAAAGCAAAT 122950

Qy        677 TTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACA 736
              |||||  |||| |||| |||||||||||||| ||| ||||||||||||| ||||| ||
Db     122949 GTATTTATGCATATACA-TTTGGCAGCCTAGGTTGTATACATGTGTACACAATCTGAGCA 122891

Qy        737 TGTGTGTATATTGAAAATCTTGTCTCTTTTTTATT 771
              | |||||||||||||||||    | ||  || ||
Db     122890 TCTGTGTATATTGAAAATCTTTCTTTTTAATTGTT 122856
```

RESULT 7

```
AC135179
LOCUS       AC135179               65160 bp    DNA     linear   HTG 10-OCT-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-1200B1 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC135179
VERSION     AC135179.2  GI:23683246
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 65160)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-1200B1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65160)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 65160)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 10, 2002 this sequence version replaced gi:23592138.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28344
                Center clone name: 1200_B_1
            --------------.
            * NOTE: This record contains 80 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      720: contig of 720 bp in length
            *      721      820: gap of 100 bp
            *      821     1524: contig of 704 bp in length
            *     1525     1624: gap of 100 bp
            *     1625     2325: contig of 701 bp in length
            *     2326     2425: gap of 100 bp
            *     2426     3211: contig of 786 bp in length
            *     3212     3311: gap of 100 bp
            *     3312     4038: contig of 727 bp in length
            *     4039     4138: gap of 100 bp
            *     4139     4869: contig of 731 bp in length
            *     4870     4969: gap of 100 bp
            *     4970     5704: contig of 735 bp in length
            *     5705     5804: gap of 100 bp
            *     5805     6538: contig of 734 bp in length
            *     6539     6638: gap of 100 bp
            *     6639     7365: contig of 727 bp in length
            *     7366     7465: gap of 100 bp
            *     7466     8184: contig of 719 bp in length
            *     8185     8284: gap of 100 bp
            *     8285     8988: contig of 704 bp in length
            *     8989     9088: gap of 100 bp
            *     9089     9806: contig of 718 bp in length
            *     9807     9906: gap of 100 bp
            *     9907    10615: contig of 709 bp in length
            *    10616    10715: gap of 100 bp
            *    10716    11447: contig of 732 bp in length
            *    11448    11547: gap of 100 bp
            *    11548    12264: contig of 717 bp in length
            *    12265    12364: gap of 100 bp
            *    12365    13073: contig of 709 bp in length
            *    13074    13173: gap of 100 bp
            *    13174    13891: contig of 718 bp in length
            *    13892    13991: gap of 100 bp
            *    13992    14714: contig of 723 bp in length
            *    14715    14814: gap of 100 bp
            *    14815    15516: contig of 702 bp in length
            *    15517    15616: gap of 100 bp
            *    15617    16335: contig of 719 bp in length
            *    16336    16435: gap of 100 bp
            *    16436    17152: contig of 717 bp in length
            *    17153    17252: gap of 100 bp
            *    17253    17967: contig of 715 bp in length
            *    17968    18067: gap of 100 bp
            *    18068    18788: contig of 721 bp in length
            *    18789    18888: gap of 100 bp
            *    18889    19629: contig of 741 bp in length
            *    19630    19729: gap of 100 bp
            *    19730    20457: contig of 728 bp in length
            *    20458    20557: gap of 100 bp
            *    20558    21286: contig of 729 bp in length
            *    21287    21386: gap of 100 bp
            *    21387    22110: contig of 724 bp in length
            *    22111    22210: gap of 100 bp
            *    22211    22922: contig of 712 bp in length
            *    22923    23022: gap of 100 bp
            *    23023    23744: contig of 722 bp in length
            *    23745    23844: gap of 100 bp
            *    23845    24575: contig of 731 bp in length
            *    24576    24675: gap of 100 bp
            *    24676    25397: contig of 722 bp in length
            *    25398    25497: gap of 100 bp
            *    25498    26104: contig of 607 bp in length
```

```
            *    26105    26204: gap of 100 bp
            *    26205    26914: contig of 710 bp in length
            *    26915    27014: gap of 100 bp
            *    27015    27746: contig of 732 bp in length
            *    27747    27846: gap of 100 bp
            *    27847    28566: contig of 720 bp in length
            *    28567    28666: gap of 100 bp
            *    28667    29399: contig of 733 bp in length
            *    29400    29499: gap of 100 bp
            *    29500    30225: contig of 726 bp in length
            *    30226    30325: gap of 100 bp
            *    30326    31052: contig of 727 bp in length
            *    31053    31152: gap of 100 bp
            *    31153    31871: contig of 719 bp in length
            *    31872    31971: gap of 100 bp
            *    31972    32686: contig of 715 bp in length
            *    32687    32786: gap of 100 bp
            *    32787    33508: contig of 722 bp in length
            *    33509    33608: gap of 100 bp
            *    33609    34328: contig of 720 bp in length
            *    34329    34428: gap of 100 bp
            *    34429    35148: contig of 720 bp in length
            *    35149    35248: gap of 100 bp
            *    35249    35958: contig of 710 bp in length
            *    35959    36058: gap of 100 bp
            *    36059    36750: contig of 692 bp in length
            *    36751    36850: gap of 100 bp
            *    36851    37555: contig of 705 bp in length
            *    37556    37655: gap of 100 bp
            *    37656    38394: contig of 739 bp in length
            *    38395    38494: gap of 100 bp
            *    38495    39232: contig of 738 bp in length
            *    39233    39332: gap of 100 bp
            *    39333    40060: contig of 728 bp in length
            *    40061    40160: gap of 100 bp
            *    40161    40878: contig of 718 bp in length
            *    40879    40978: gap of 100 bp
            *    40979    41703: contig of 725 bp in length
            *    41704    41803: gap of 100 bp
            *    41804    42462: contig of 659 bp in length
            *    42463    42562: gap of 100 bp
            *    42563    43291: contig of 729 bp in length
            *    43292    43391: gap of 100 bp
            *    43392    44102: contig of 711 bp in length
            *    44103    44202: gap of 100 bp
            *    44203    44934: contig of 732 bp in length
            *    44935    45034: gap of 100 bp
            *    45035    45722: contig of 688 bp in length
            *    45723    45822: gap of 100 bp
            *    45823    46553: contig of 731 bp in length
            *    46554    46653: gap of 100 bp
            *    46654    47365: contig of 712 bp in length
            *    47366    47465: gap of 100 bp
            *    47466    48194: contig of 729 bp in length
            *    48195    48294: gap of 100 bp

  Query Match             54.2%;  Score 421;  DB 2;  Length 65160;
  Best Local Similarity   96.6%;  Pred. No. 2.2e-83;
  Matches  430;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        333 GGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTT 392
              |||    |||||||||||||||  || ||||||||||||||||||||||||||||||||||||
Db      33633 GGAATTCTTTTGCCATGTATAACTTCTCAAACATTAACTTTCAGGGTTATTAATCCTTTT 33692

Qy        393 AAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAA 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33693 AAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAA 33752

Qy        453 AAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATT 512
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      33753 AAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTACGTGAAGTGATT 33812

Qy        513 TTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTA 572
              |||||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||||
Db      33813 CTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTCTCCATCCCTGCATGATTTA 33872

Qy        573 TGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCA 632
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      33873 TGATGAATCATTTTGTGGTCTGTTAGTTACTTCTAGAGAATAGAAAGCATTGTAGGCTCA 33932

Qy        633 GGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTAC 692
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33933 TGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTAC 33992

Qy        693 ATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAA 752
              ||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Db      33993 ATTTTGGCAGCCTAAGCTGTGTACATGTGTACACATTCTGAACATGTGCGTATATTGAAA 34052

Qy        753 ATCTTGTCTCTTTTTTATTGTTAAG 777
              |||||||||||||||||||| ||||
Db      34053 ATCTTGTCTCTTTTTTATTGCTAAG 34077
```

```
RESULT 8
AC131898
LOCUS       AC131898              212216 bp    DNA     linear   HTG 27-AUG-2002
DEFINITION  Oryctolagus cuniculus clone LB1-83M7, WORKING DRAFT SEQUENCE, 12
            unordered pieces.
ACCESSION   AC131898
VERSION     AC131898.1  GI:22507060
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Oryctolagus cuniculus (rabbit)
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 212216)
  AUTHORS   Martin,J., Schwartz,J.R., Hosseini,R., Peng,Y., Peng,Z., Rubin,E.M.
            and Cheng,J.-F.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212216)
  AUTHORS   Martin,J., Schwartz,J.R., Hosseini,R., Peng,Y., Peng,Z., Rubin,E.M.
            and Cheng,J.-F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2002) Genome Sciences, Lawrence Berkeley National
            Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
COMMENT     Draft Sequence Produced by Berkeley PGA
            Web site: http://pga.lbl.gov
            Center Code: PGABERK
            Center Project Name: R052-83M7
            Bac Clone Name: LB1-83M7

            Additional information on comparative analysis and ordering are
            available at:
            http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=PPARG
            Funding agent: Programs for Genomic Applications (NHLBI)
            Contact: 'Jody Schwartz' jrschwartz@lbl.gov
            if library name is LB1 to LB4, please see website
            for a description: http://www-gsd.lbl.gov/cheng/BAC.html
            Summary Statistics:
            Sequencing vector: Plasmid; pUC18
            Chemistry: Dye-terminator Big Dye
            Assembly program: Phrap version 0.990329.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2430: contig of 2430 bp in length
            *     2431     2530: gap of unknown length
            *     2531     4596: contig of 2066 bp in length
            *     4597     4696: gap of unknown length
            *     4697     9216: contig of 4520 bp in length
```

```
            *     9217     9316: gap of unknown length
            *     9317    15539: contig of 6223 bp in length
            *    15540    15639: gap of unknown length
            *    15640    21088: contig of 5449 bp in length
            *    21089    21188: gap of unknown length
            *    21189    33983: contig of 12795 bp in length
            *    33984    34083: gap of unknown length
            *    34084    48093: contig of 14010 bp in length
            *    48094    48193: gap of unknown length
            *    48194    67641: contig of 19448 bp in length
            *    67642    67741: gap of unknown length
            *    67742    91391: contig of 23650 bp in length
            *    91392    91491: gap of unknown length
            *    91492   117177: contig of 25686 bp in length
            *   117178   117277: gap of unknown length
            *   117278   144157: contig of 26880 bp in length
            *   144158   144257: gap of unknown length
            *   144258   212216: contig of 67959 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212216
                     /organism="Oryctolagus cuniculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9986"
                     /clone="LB1-83M7"
ORIGIN

  Query Match           16.8%;  Score 130.6;  DB 2;  Length 212216;
  Best Local Similarity   62.5%;  Pred. No. 4.8e-19;
  Matches  307;  Conservative    0;  Mismatches  144;  Indels   40;  Gaps    5;

Qy        287 GGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCC 346
              ||||||| |||||  || |  || ||||||| ||||| || |  ||  ||  | | || |
Db      84838 GGCTTAACAACCACCTGCCGCCTGGAATGAATTTATA-TTCAGGAACTATATTCTATGAC 84896

Qy        347 ATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTT 406
              | | |  ||  ||      |  |||||  | |||||||||||||||||| ||||||||||
Db      84897 TTCTGTTGATCATC------TGCCTTTCCAGATTATTAATCCTTTTAAGATCTAGTTTTT 84950

Qy        407 CTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGG 466
              ||||||||||    | ||||| | |||||    ||| || || ||||||||||||    |
Db      84951 CTTAAGTCTGCATAATAATATATGTATCAGTATTAATTTGGCATAAAAGATGGATGAGAT 85010

Qy        467 CTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTC 526
                | ||       |   |   || | ||  ||        ||  |   |
Db      85011 AAAGGACCTTGGACTCACATTAGTCACTGAACTGGGAAAATAAGCGAATCTAATTCATGT 85070

Qy        527 TTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTT 586
              |    |||| ||||| | |||||| | || |||||  |||||||||||||| ||| |
Db      85071 CTCATAAAAATCATTTTAAGAAGATTTTGCATCCCTGTGTGATTTATGATGAGTCACTGG 85130

Qy        587 GTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACAT 646
              |||| ||||| ||||||| ||||||| | ||||                      |||
Db      85131 GTGGCCTGTTGGTTACTTCTAGAGAACAAAAAG----------------------CAT 85166

Qy        647 TCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTA 706
              |||||||| ||||||| | || |   |||||||||| |  | |  ||||||||| ||||
Db      85167 TCAGAATGGAATCCAACAAAGGA-AAAAATTTATTTGTGTGTATGTATTTTGGCAACCTA 85225

Qy        707 GGCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTT 766
              |          ||| ||   ||| | || ||||||||||||||||| || || | ||||||
Db      85226 G--------ATTGTCTAAGGATTTTAAATATGTGTGTATATTGAACATTTTATTTCTTTT 85277

Qy        767 TTATTGTTAAG 777
                ||||||  ||
Db      85278 GAATTGTTCAG 85288


RESULT 9
AC125447
LOCUS       AC125447              185737 bp    DNA     linear   HTG 06-MAR-2004
DEFINITION  Mus musculus chromosome 6 clone RP24-507D15 map 6, *** SEQUENCING
            IN PROGRESS ***, 3 unordered pieces.
ACCESSION   AC125447
VERSION     AC125447.4  GI:45237223
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 185737)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 6, clone RP24-507D15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185737)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 185737)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 6, 2004 this sequence version replaced gi:29029334.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26714
                Center clone name: 507_D_15
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
```

```
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    54649: contig of 54649 bp in length
            *    54650    54749: gap of 100 bp
            *    54750   156594: contig of 101845 bp in length
            *   156595   156694: gap of 100 bp
            *   156695   185737: contig of 29043 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185737
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /map="6"
                     /clone="RP24-507D15"
                     /clone_lib="RPCI-24 Male Mouse BAC"
ORIGIN

  Query Match             13.3%;  Score 103.6;  DB 2;  Length 185737;
  Best Local Similarity   57.7%;  Pred. No. 4.8e-13;
  Matches  225;  Conservative    0;  Mismatches  159;  Indels    6;  Gaps    2;

Qy        371 TTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGG 430
              ||| ||| |  | |  ||||| ||  ||||| ||||| |||||||||    | | ||| |
Db     129614 TTTGAGGTTGCTAATCCCTTTAAAAATCTAGGTTTTCATAAGTCTGAATACTGATATAAT 129673

Qy        431 TATCGTCA--TTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGA 488
              |||  | |   | ||   | | |   |     | || |  ||   ||||| |||||
Db     129674 TATTATTAACCTGATACTAAAAATGGGTATAATAGAGTTTTGCATCATGTCAGTGACCCT 129733

Qy        489 AATAGGAAAGTAGGTGAAGTGATTTTAATAGATGT----TTCTTTTATGAAATAATTTTT 544
              |   | |||  | | ||||||  |  || || | |    |||||||||| ||  | | |
Db     129734 ACAGGTAAAACAAGAGAAGTGGATCAAACAGCTATCTGCTTCTTTTATGCAACTACTCTG 129793

Qy        545 AAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTT 604
              |  ||| | |   |  |||  ||||  ||||||| || |||  |||| | || |||| ||
Db     129794 ATGAGAGTTTTGGGTACTGTGTGATCCATGATGAGTCGTTTAATGGTGTATTGGTTATTT 129853

Qy        605 TTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATA 664
              |||||||||  ||||  || || |||||||||||||||| | |  | ||||||   |||
Db     129854 CTAGAGAATATAAAGTTTTATATGCTCAGGGAAAGCAAATACTGGGGATGAAAATTAATG 129913

Qy        665 GAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTAC 724
              |  |   || |   |  || |   | ||  |   |||   |  ||| | | |
Db     129914 TGTATATTTGCATTTTGGCAGATTGGAATGTCTGAGAATTTAGAAAATCAACAAATTTTC 129973

Qy        725 ACATTCTGAACATGTGTGTATATTGAAAAT 754
              | | | |     | | |   |||| | | |
Db     129974 AAAATATTTTTCTTTTTAATTATTCAGATT 130003


RESULT 10
AC136055/c
LOCUS       AC136055              191540 bp    DNA     linear   HTG 29-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-123L10, *** SEQUENCING IN PROGRESS
            ***, 62 unordered pieces.
ACCESSION   AC136055
VERSION     AC136055.1  GI:24417926
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 191540)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191540)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KDDG
                Center clone name: CH230-123L10
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 138097 bases at least Q40
                Consensus quality: 143025 bases at least Q30
                Consensus quality: 147154 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 62 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1135: contig of 1135 bp in length
*     1136     1235: gap of unknown length
*     1236     2684: contig of 1449 bp in length
*     2685     2784: gap of unknown length
*     2785     4087: contig of 1303 bp in length
*     4088     4187: gap of unknown length
*     4188     5601: contig of 1414 bp in length
*     5602     5701: gap of unknown length
*     5702     7184: contig of 1483 bp in length
*     7185     7284: gap of unknown length
*     7285     8589: contig of 1305 bp in length
*     8590     8689: gap of unknown length
*     8690    10203: contig of 1514 bp in length
*    10204    10303: gap of unknown length
*    10304    11907: contig of 1604 bp in length
*    11908    12007: gap of unknown length
*    12008    13032: contig of 1025 bp in length
*    13033    13132: gap of unknown length
*    13133    14692: contig of 1560 bp in length
*    14693    14792: gap of unknown length
*    14793    15997: contig of 1205 bp in length
*    15998    16097: gap of unknown length
*    16098    17167: contig of 1070 bp in length
*    17168    17267: gap of unknown length
*    17268    18690: contig of 1423 bp in length
*    18691    18790: gap of unknown length
*    18791    20652: contig of 1862 bp in length
*    20653    20752: gap of unknown length
*    20753    22248: contig of 1496 bp in length
*    22249    22348: gap of unknown length
*    22349    24000: contig of 1652 bp in length
*    24001    24100: gap of unknown length
*    24101    25384: contig of 1284 bp in length
*    25385    25484: gap of unknown length
*    25485    26797: contig of 1313 bp in length
*    26798    26897: gap of unknown length
*    26898    28389: contig of 1492 bp in length
*    28390    28489: gap of unknown length
*    28490    29812: contig of 1323 bp in length
*    29813    29912: gap of unknown length
*    29913    31236: contig of 1324 bp in length
*    31237    31336: gap of unknown length
*    31337    33116: contig of 1780 bp in length
*    33117    33216: gap of unknown length
*    33217    34328: contig of 1112 bp in length
*    34329    34428: gap of unknown length
*    34429    37011: contig of 2583 bp in length
*    37012    37111: gap of unknown length
*    37112    39760: contig of 2649 bp in length
*    39761    39860: gap of unknown length
*    39861    41317: contig of 1457 bp in length
*    41318    41417: gap of unknown length
*    41418    44017: contig of 2600 bp in length
*    44018    44117: gap of unknown length
*    44118    45721: contig of 1604 bp in length
*    45722    45821: gap of unknown length
*    45822    47208: contig of 1387 bp in length
*    47209    47308: gap of unknown length
*    47309    49161: contig of 1853 bp in length
*    49162    49261: gap of unknown length
*    49262    52550: contig of 3289 bp in length
*    52551    52650: gap of unknown length
*    52651    54753: contig of 2103 bp in length
*    54754    54853: gap of unknown length
*    54854    56484: contig of 1631 bp in length
*    56485    56584: gap of unknown length
*    56585    58264: contig of 1680 bp in length
*    58265    58364: gap of unknown length
*    58365    61440: contig of 3076 bp in length
*    61441    61540: gap of unknown length
*    61541    63616: contig of 2076 bp in length
*    63617    63716: gap of unknown length
*    63717    65642: contig of 1926 bp in length
*    65643    65742: gap of unknown length
*    65743    68360: contig of 2618 bp in length
*    68361    68460: gap of unknown length
*    68461    70926: contig of 2466 bp in length
*    70927    71026: gap of unknown length
*    71027    74092: contig of 3066 bp in length
*    74093    74192: gap of unknown length
*    74193    76455: contig of 2263 bp in length
*    76456    76555: gap of unknown length
*    76556    79002: contig of 2447 bp in length
*    79003    79102: gap of unknown length
*    79103    82482: contig of 3380 bp in length
*    82483    82582: gap of unknown length
*    82583    85503: contig of 2921 bp in length
*    85504    85603: gap of unknown length
*    85604    87769: contig of 2166 bp in length
*    87770    87869: gap of unknown length
*    87870    90671: contig of 2802 bp in length
*    90672    90771: gap of unknown length
*    90772    94059: contig of 3288 bp in length
*    94060    94159: gap of unknown length
*    94160    98678: contig of 4519 bp in length
*    98679    98778: gap of unknown length
*    98779   104079: contig of 5301 bp in length
*   104080   104179: gap of unknown length
*   104180   108047: contig of 3868 bp in length
*   108048   108147: gap of unknown length
*   108148   112386: contig of 4239 bp in length
*   112387   112486: gap of unknown length
*   112487   116184: contig of 3698 bp in length
*   116185   116284: gap of unknown length
*   116285   122755: contig of 6471 bp in length
*   122756   122855: gap of unknown length

  Query Match             11.4%;  Score 88.6;  DB 2;  Length 191540;
  Best Local Similarity   60.7%;  Pred. No. 1e-09;
  Matches  184;  Conservative    0;  Mismatches  109;  Indels   10;  Gaps    2;

Qy        371 TTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGG 430
              ||| ||  |  | |  ||||| ||| ||||| ||||| ||||||||    |||||  |
Db      13980 TTTGAGCTTGCTAATCCCTTTAAAGATCTAGGTTTTCCTAAGTCTGAATGGTAATGTAAT 13921

Qy        431 TATCGTCATTCATGTGACATAAAAG------ATGGAAAGGGGCTTCATTCATGTTAGTGA 484
              |||  | | |    |  | ||||      |    ||||    ||   |||||| |||||
Db      13920 TATTATTAATTTGATATTAAAAAATGGACATAATAAAAGACTTTTGCATCATGTCAGTGA 13861

Qy        485 TGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGA----TGTTTCTTTTATGAAATAAT 540
                | | | |||  | | |||||||  |  ||||      ||| |||||||||  ||  ||
Db      13860 CCCTAGTGGTAAAACAAGAGAAGTGAATCAAATACCTATCTGTCTCTTTTATGCAACCAT 13801

Qy        541 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600
              | | |  ||| | |      |             | ||||| |||||| ||||| |||| ||
Db      13800 TCTGAGGAGAGTTTGGGTGCTGCATGATCCAACGATGAGTCATTTAGTGGTGTGTTGGTG 13741

Qy        601 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660
              | || |||||||| ||||| ||  | |||| | |||||||| | |  | ||||||
Db      13740 ATTTCTAGAGAATATAAAGCTTTACATGCTCGGAGAAAGCAAATACTGTGGATGAAAATT 13681

Qy        661 AAT 663
              |||
Db      13680 AAT 13678


RESULT 11
AC091418/c
LOCUS       AC091418              245724 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-2G6, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC091418
VERSION     AC091418.5  GI:24941354
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
```

```
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 245724)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245724)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 245724)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:22855703.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: TUQY
                Center clone name: CH230-2G6
            ------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 220652 bases at least Q40
                Consensus quality: 223895 bases at least Q30
                Consensus quality: 225929 bases at least Q20
                Estimated insert size: 229375; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   245724: contig of 245724 bp in length.
FEATURES            Location/Qualifiers
     source         1. .245724
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-2G6"
     misc_feature   1. .1001
                    /note="wgs_contig"
     misc_feature   243338. .245724
                    /note="wgs_contig"
ORIGIN

  Query Match            11.4%;  Score 88.6;  DB 2;  Length 245724;
  Best Local Similarity   60.7%;  Pred. No. 9.8e-10;
  Matches  184;  Conservative    0;  Mismatches  109;  Indels   10;  Gaps    2;

Qy        371 TTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGG 430
              ||| ||  |  | |  ||||| ||| ||||| ||||| ||||||||     |||||  |
Db      52722 TTTGAGCTTGCTAATCCCTTTAAAGATCTAGGTTTTCCTAAGTCTGAATGGTAATGTAAT 52663

Qy        431 TATCGTCATTCATGTGACATAAAAG------ATGGAAAGGGGCTTCATTCATGTTAGTGA 484
              |||  | | |     |   | ||||       |   ||||    ||   |||||| |||||
Db      52662 TATTATTAATTTGATATTAAAAAATGGACATAATAAAAGACTTTTGCATCATGTCAGTGA 52603

Qy        485 TGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGA----TGTTTCTTTTATGAAATAAT 540
                | | |||  | | |||||||  |  ||||      ||| |||||||||| ||  ||
Db      52602 CCCTAGTGGTAAAACAAGAGAAGTGAATCAAATACCTATCTGTCTCTTTTATGCAACCAT 52543

Qy        541 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600
              | | |  ||| | |      |           | ||||| |||||| ||||| |||| ||
Db      52542 TCTGAGGAGAGTTTGGGTGCTGCATGATCCAACGATGAGTCATTTAGTGGTGTGTTGGTG 52483

Qy        601 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660
              | || |||||||||| ||||| ||   | |||| | |||||||||| | |  | ||||||
Db      52482 ATTTCTAGAGAATATAAAGCTTTACATGCTCGGAGAAAGCAAATACTGTGGATGAAAATT 52423

Qy        661 AAT 663
              |||
Db      52422 AAT 52420


RESULT 12
AC120668/c
LOCUS       AC120668              250169 bp    DNA     linear   HTG 23-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-24K10, *** SEQUENCING IN PROGRESS
```

```
            ***, 3 unordered pieces.
ACCESSION   AC120668
VERSION     AC120668.4  GI:25188255
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 250169)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,P., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 250169)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 250169)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 23, 2002 this sequence version replaced gi:23322270.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GXLO
                Center clone name: CH230-24K10
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 225414 bases at least Q40
                Consensus quality: 228258 bases at least Q30
                Consensus quality: 229994 bases at least Q20
                Estimated insert size: 226717; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9799: contig of 9799 bp in length
            *     9800     9899: gap of unknown length
            *     9900   248957: contig of 239058 bp in length
            *   248958   249057: gap of unknown length
            *   249058   250169: contig of 1112 bp in length.
FEATURES             Location/Qualifiers
     source          1. .250169
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-24K10"
     misc_feature    4754. .5526
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH280541"
     misc_feature    9900. .10968
                     /note="wgs_contig"
     misc_feature    11406. .14106
                     /note="wgs_contig"
     misc_feature    complement(234439. .235077)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH280540"
     misc_feature    245037. .246234
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    247156. .248957
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN

  Query Match             11.4%;  Score 88.6;  DB 2;  Length 250169;
```

```
  Best Local Similarity   60.7%;  Pred. No. 9.8e-10;
  Matches  184;  Conservative    0;  Mismatches  109;  Indels   10;  Gaps    2;

Qy        371 TTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGG 430
              ||| ||  |  | |  ||||| ||| ||||| ||||| ||||||||     |||||  |
Db     191916 TTTGAGCTTGCTAATCCCTTTAAAGATCTAGGTTTTCCTAAGTCTGAATGGTAATGTAAT 191857

Qy        431 TATCGTCATTCATGTGACATAAAAG------ATGGAAAGGGGCTTCATTCATGTTAGTGA 484
              |||  | | |     |   | ||||       |   ||||    ||   |||||| |||||
Db     191856 TATTATTAATTTGATATTAAAAAATGGACATAATAAAAGACTTTTGCATCATGTCAGTGA 191797

Qy        485 TGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGA----TGTTTCTTTTATGAAATAAT 540
                | | | |||  | | ||||||| |  ||||     ||| ||||||||| ||  ||
Db     191796 CCCTAGTGGTAAAACAAGAGAAGTGAATCAAATACCTATCTGTCTCTTTTATGCAACCAT 191737

Qy        541 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600
              | | |  ||| | |     |           | ||||| |||||| ||||| |||| ||
Db     191736 TCTGAGGAGAGTTTGGGTGCTGCATGATCCAACGATGAGTCATTTAGTGGTGTGTTGGTG 191677

Qy        601 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660
              | || |||||||| ||||| ||  | |||| | |||||||| | |  | ||||||
Db     191676 ATTTCTAGAGAATATAAAGCTTTACATGCTCGGAGAAAGCAAATACTGTGGATGAAAATT 191617

Qy        661 AAT 663
              |||
Db     191616 AAT 191614


RESULT 13
AC142242
LOCUS       AC142242               154195 bp    DNA     linear   HTG 07-AUG-2003
DEFINITION  Atelerix albiventris clone LB4-81H3, WORKING DRAFT SEQUENCE, 5
            ordered pieces.
ACCESSION   AC142242
VERSION     AC142242.2  GI:33469169
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Atelerix albiventris (middle-African hedgehog)
  ORGANISM  Atelerix albiventris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Insectivora; Erinaceidae; Erinaceinae;
            Atelerix.
REFERENCE   1  (bases 1 to 154195)
  AUTHORS   Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
            Peng,Z., Malinov,I. and Rubin,E.M.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154195)
  AUTHORS   Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
            Peng,Z., Malinov,I. and Rubin,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Genome Sciences, Lawrence Berkeley National
            Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
REFERENCE   3  (bases 1 to 154195)
  AUTHORS   Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
            Peng,Z., Malinov,I. and Rubin,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-2003) Genome Sciences, Lawrence Berkeley National
            Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
COMMENT     On Aug 7, 2003 this sequence version replaced gi:29294044.

            Sequence Produced by Berkeley PGA
            Web site: http://pga.lbl.gov
            Center Code: PGABERK
            Center Project Name: E015
            Bac Clone Name: LB4-81H3

            This sequence has been compared to sequences of other species
            using Vista (http://www-gsd.lbl.gov/VISTA). The results can be
            viewed at:
            http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=PPARG

            The order-orientation of the draft sequence was accomplished by
            using:
            Avid (http://baboon.math.berkeley.edu/mavid),
            Lagan (http://lagan.stanford.edu/) and paired end information.

            Funding agent: Programs for Genomic Applications (NHLBI)

            If the Bac Library Name is LB1 to LB4, please see website
            for the description: http://www-gsd.lbl.gov/cheng/BAC.html
            These libraries are available through the BACPAC Resources Center:
            http://www.chori.org/bacpac/libraryres.htm as LBNL-1 to LBNL-4.

            Summary Statistics:
            Sequencing vector: Plasmid; pUC18
            Chemistry: Dye-terminator Big Dye
            Assembly program: Phrap version 0.990329.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    33275: contig of 33275 bp in length
            *    33276    33375: gap of unknown length
            *    33376    54251: contig of 20876 bp in length
            *    54252    54351: gap of unknown length
            *    54352    65733: contig of 11382 bp in length
            *    65734    65833: gap of unknown length
            *    65834   127515: contig of 61682 bp in length
            *   127516   127615: gap of unknown length
            *   127616   154195: contig of 26580 bp in length.
FEATURES             Location/Qualifiers
     source          1. .154195
                     /organism="Atelerix albiventris"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9368"
                     /clone="LB4-81H3"
ORIGIN

  Query Match             8.8%;  Score 68.6;  DB 2;  Length 154195;
  Best Local Similarity   62.3%;  Pred. No. 2.9e-05;
  Matches  124;  Conservative    0;  Mismatches   74;  Indels    1;  Gaps    1;

Qy        558 GCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAA 617
              | |||| | |||| |  | || |||  |  | ||||||||||| ||| | | ||| | ||
Db      17762 GTCCTGTAAGATTGACCAGGAGTCAACTGCTAGTCTGTTAGTTGCTTCTGGGGAAAACAA 17821

Qy        618 AGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATT 677
              |||||  ||  | |||||  |   |  || ||||  |||| | ||  | ||||   ||| |
Db      17822 AGCATGATAAACACAGGGGACAGAG-CACTCAGTGTGAAGTACAGCACAGAAAACAAAGT 17880

Qy        678 TATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACAT 737
              || ||| || || ||||||||||||||| ||  || ||    | |  |    |   ||
Db      17881 TAGTTGTGCTTGAACATTTTGGCAGCCTTGGTGGTCTAAGACTCTGGAATGTTAAAAATG 17940

Qy        738 GTGTGTATATTGAAAATCT 756
              ||| | | ||  |  | |
Db      17941 TTGTTTCTTTTTTACTTTT 17959


RESULT 14
BX465190
LOCUS       BX465190               183103 bp    DNA     linear   VRT 13-NOV-2003
DEFINITION  Zebrafish DNA sequence from clone DKEY-121A11 in linkage group 22,
            complete sequence.
ACCESSION   BX465190
VERSION     BX465190.4  GI:38323074
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
```

```
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 183103)
  AUTHORS   Barlow,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 13, 2003 this sequence version replaced gi:32567584.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-121A11
            is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1. .183103
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-121A11"
                     /clone_lib="DanioKey"
ORIGIN

  Query Match             6.8%;  Score 53;  DB 5;  Length 183103;
  Best Local Similarity   48.5%;  Pred. No. 0.081;
  Matches  176;  Conservative    0;  Mismatches  185;  Indels    2;  Gaps    1;

Qy        321 ATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTT 380
              |||| || ||    |   ||       |||||| ||||||||| |||      ||   |
Db     159530 ATATGTTTAATGCTACTATTACCTGACTGTATACATTTTCAAATATTTTACAACACATTG 159589

Qy        381 ATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGT--ATCGTCA 438
              || |  | |  ||| |    |   | ||  || | |||  | ||    ||  ||| |
Db     159590 TTTTAAACCTAAAAGTTTCTATAACTATTGTGTTTTTGCCCTGATTACAGTACATCATAT 159649

Qy        439 TTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAG 498
              || |   |  ||   ||| |  |   |  ||||         |  || ||||| |  ||
Db     159650 TTTACTAGTTATTTTAGAAGATACTAGTATTCAGCTTAAAGTGGAATTGAAATGGTTAAT 159709

Qy        499 TAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAG 558
              |||||||||| |||  |  | |   ||| || || ||  | |   |||| ||  |  |
Db     159710 TAGGTGAAGTCATTGGACAACACTGGTCTGTTGTGCAACCAATCGAAAAAAATAATTGAT 159769

Qy        559 CCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAA 618
              |  |||      |  || | |       ||| ||  ||  | |  |||
Db     159770 AATATTGACCATATTGACCTGGAAAATTATTTTAAAAACTTTAATTACAGTCAGTCTAAA 159829

Qy        619 GCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTT 678
              |   || || |  | |  ||     |||||  |||  || || | |  |||||||
Db     159830 AGAAATAAGACTGAAATAGAAAAATAGAACAGAAGAAAAAACATTAAATACCGTAAATTC 159889

Qy        679 ATT 681
              |||
Db     159890 CTT 159892


RESULT 15
CR847935
LOCUS       CR847935              145160 bp    DNA     linear   HTG 11-OCT-2004
DEFINITION  Danio rerio clone CH211-194N10, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR847935
VERSION     CR847935.2  GI:54034952
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 145160)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 11, 2004 this sequence version replaced gi:54019745.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC194N10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 141804 bases at least Q40
            Consensus quality: 142048 bases at least Q30
            Consensus quality: 142375 bases at least Q20
            Insert size: 144560; sum-of-contigs
            Insert size: 150827; 3.5% error; agarose-fp
            Quality coverage: 11.77x in Q20 bases; sum-of-contigs Quality
            coverage: 11.28x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    47977: contig of 47977 bp in length
            *    47978    48077: gap of 100 bp
            *    48078    61854: contig of 13777 bp in length
            *    61855    61954: gap of 100 bp
            *    61955   114748: contig of 52794 bp in length
            *   114749   114848: gap of 100 bp
            *   114849   119330: contig of 4482 bp in length
            *   119331   119430: gap of 100 bp
            *   119431   127887: contig of 8457 bp in length
            *   127888   127987: gap of 100 bp
            *   127988   134017: contig of 6030 bp in length
            *   134018   134117: gap of 100 bp
            *   134118   145160: contig of 11043 bp in length.
FEATURES             Location/Qualifiers
```

```
     source          1. .145160
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-194N10"
                     /clone_lib="CHORI-211"
     misc_feature    1. .47977
                     /note="assembly_fragment:00776
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    48078. .61854
                     /note="assembly_fragment:00558
                     fragment_chain:1"
     misc_feature    61955. .114748
                     /note="assembly_fragment:01633
                     fragment_chain:2"
     misc_feature    114849. .119330
                     /note="assembly_fragment:00004
                     fragment_chain:2"
     misc_feature    119431. .127887
                     /note="assembly_fragment:00176
                     fragment_chain:2"
     misc_feature    127988. .134017
                     /note="assembly_fragment:00071
                     fragment_chain:2"
     misc_feature    134118. .145160
                     /note="assembly_fragment:00347
                     fragment_chain:2
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             6.8%;  Score 52.8;  DB 2;  Length 145160;
  Best Local Similarity   44.9%;  Pred. No. 0.093;
  Matches  201;  Conservative    0;  Mismatches  247;  Indels    0;  Gaps    0;

Qy        320 TATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGT 379
              | |||||     |   ||   |||     || |||  ||||   || | |||   | |   |
Db       9856 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAATAATAATAATAATTATTAT 9915

Qy        380 TATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCAT 439
              ||||| |  | |||   |     || || | | | |  | |||| ||    |||    |
Db       9916 TATTATTATTATTATTATTATTATTATCATTATTATTAGTAGTATTATCATTATTAGTAG 9975

Qy        440 TCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGT 499
              | | | |      |  ||     |       | |||  | ||| |  |    | |    |  |
Db       9976 TATTATTATTAGTATTATTATCATTATTATTATTATTATTATTATTTACACTTTCTATTT 10035

Qy        500 AGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGC 559
              |       || | ||| ||    | |  ||  |   || ||| |||  |   |    |
Db      10036 ACTGTTGCTGCTGTTATTATTAATATTATTATTATTATTATTATTATTATTATTATT 10095

Qy        560 CCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAG 619
              |      ||||    ||   |||| | ||||  |     |    ||  ||   | |
Db      10096 ATTATTATTATTATTTATAAATGTTTTATTGTCTTGTTACGAACGTTTAAGTGGAAAGTC 10155

Qy        620 CATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTTA 679
              |  |    || |  | |  || |  |   | | ||    |||| | ||  |||||  |
Db      10156 TTTCATTATGTCTGTCATATTAAGCTATTTCACTTAATAAAAATAAACAAAATAAATAAA 10215

Qy        680 TTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACATGT 739
              |       |  || ||   |     | |  | |   ||   |   | ||  ||
Db      10216 TAAAACTTTAAATATGTTAAGAAATATGACCCTTTCTCTGCACAGCAAGTCAAAAATAAA 10275

Qy        740 GTGTATATTGAAAATCTTGTCTCTTTTT 767
              |||||||||  ||  |||||| |||||
Db      10276 TTGTATATTGTCAAAATTGTCTTTTTTT 10303


Search completed: November  9, 2005, 02:15:48
Job time : 4083.01 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:46:48 ; Search time 629.342 Seconds
(without alignments)
7308.644 Million cell updates/sec

Title: US-09-463-542-34_COPY_368_1144
Perfect score: 777
Sequence: 1 tcatgtaggtaagactgtgt.....gtctctttttattgttaag 777

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	777	100.0	1463	2	AAX19034	Aax19034 Human PPA
2	764.2	98.4	158417	13	ADS36461	Ads36461 Human aut
3	383	49.3	468	2	AAX19065	Aax19065 Human PPA
4	200.6	25.8	201	13	ADS39355	Ads39355 Human aut
5	200.6	25.8	201	13	ADS39165	Ads39165 Human aut
6	196.6	25.3	201	13	ADS39169	Ads39169 Human aut
7	195.8	25.2	201	13	ADS39163	Ads39163 Human aut
8	50	6.4	5771	6	ABN80072	Abn80072 Human che
9	49.2	6.3	18683	6	ABL32312	Abl32312 Human imm
10	49.2	6.3	18683	6	ABL54333	Abl54333 Chemicall
11	49	6.3	11422	6	ABK39937	Abk39937 Human che
12	49	6.3	11422	6	ABL32219	Abl32219 Human imm
13	48.4	6.2	13123	6	ABK31423	Abk31423 Signal tr
14	48.4	6.2	13123	6	ABL54364	Abl54364 Chemicall
15	47	6.0	6880	6	ABK31322	Abk31322 Signal tr
16	47	6.0	6880	6	ABL70293	Abl70293 Chemicall
17	47	6.0	6880	6	AAS61223	Aas61223 Human gen
18	46.8	6.0	6179	4	AAS46343	Aas46343 Tumour su
19	46.8	6.0	6179	6	ABK31250	Abk31250 Signal tr
20	46.6	6.0	33053	6	ABQ67006	Abq67006 Human ang
c 21	46.4	6.0	50000	6	ABL56202	Abl56202 AmEPV gen
22	45.2	5.8	2000	6	ABZ17494	Abz17494 Arabidops
23	45.2	5.8	5718	4	AAS46464	Aas46464 Tumour su
24	45.2	5.8	5718	6	ABL33373	Abl33373 Human imm
25	44.4	5.7	5886	6	ABL34213	Abl34213 Human imm
26	44.4	5.7	5992	6	AAS61208	Aas61208 Human gen
27	44.4	5.7	16811	6	ABL33946	Abl33946 Human imm
28	44.4	5.7	18624	6	ABL33702	Abl33702 Human imm
29	44.2	5.7	5937	6	ABL34542	Abl34542 Human met
30	44.2	5.7	5937	6	ABN80150	Abn80150 Human che
31	44.2	5.7	5937	7	ADS99803	Ads99803 Bisulphit
32	44.2	5.7	6057	6	ABK31397	Abk31397 Signal tr
33	44.2	5.7	6057	6	ABL70362	Abl70362 Chemicall
34	44.2	5.7	6823	6	ABL33145	Abl33145 Human imm
35	44	5.7	13584	6	ABL32615	Abl32615 Human imm
36	43.8	5.6	17918	6	AAS61418	Aas61418 Human gen
c 37	43.6	5.6	69727	10	ACF65374	Acf65374 Photorhab
c 38	43.6	5.6	110000	10	ACF67367_35	Continuation (36 o
c 39	43.4	5.6	2000	6	ABZ16087	Abz16087 Arabidops
40	43.4	5.6	6849	6	ABL92253	Abl92253 Chemicall
41	43.4	5.6	10543	6	ABK31249	Abk31249 Signal tr
42	43.4	5.6	10543	6	ABL70206	Abl70206 Chemicall
43	43.4	5.6	10543	6	AAS61161	Aas61161 Human gen
44	43.4	5.6	34548	6	ABL70603	Abl70603 Chemicall
45	43.2	5.6	6056	6	ABL33026	Abl33026 Human imm

ALIGNMENTS

RESULT 1
AAX19034
ID AAX19034 standard; DNA; 1463 BP.
XX
AC AAX19034;
XX
DT 13-MAY-1999 (first entry)
XX
DE Human PPAR-gamma-3 proximal promoter, exon A2 and intron A2.
XX
KW Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;
KW regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;
KW lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;
KW insulin resistance; non-insulin-dependent diabetes mellitus;
KW polycystic ovary syndrome; gastrointestinal tract; Crohn's disease;
KW inflammatory bowel disease; ulcerative colitis; bowel cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9905161-A1.
XX
PD 04-FEB-1999.
XX
PF 24-JUL-1998; 98WO-US015411.
XX
PR 25-JUL-1997; 97US-0053692P.
XX
PA (LIGA-) LIGAND PHARM INC.
PA (INSP) INST PASTEUR.
XX
PI Briggs MR, Saladin RS, Auwerx J, Fajas L;
XX
DR WPI; 1999-142844/12.
XX
PT Newly isolated nucleic acid comprising a control region of a human
PT peroxisome proliferator activated receptor (PPAR) gamma gene - useful for
PT identifying modulators that are useful in treating diseases associated
PT with abnormal levels of human PPAR-gamma gene expression.
XX
PS Claim 11; Page 88; 102pp; English.
XX
CC The present invention describes an isolated, purified or enriched nucleic
CC acid comprising a control region of a human peroxisome proliferator

```
CC   activated receptor gamma (PPAR-gamma) gene. The nucleic acids are useful
CC   for screening for agents capable of modulating the expression of a human
CC   PPAR-gamma gene. These agents (modulators) form pharmaceutical
CC   compositions that are useful for treating diseases associated with
CC   high/low levels of human PPAR-gamma gene expression. The diseases include
CC   obesity, anorexia, cachexia, lipodystrophy, lipomas, liposarcomas,
CC   abnormalities associated with anti-human immunodeficiency virus (HIV)
CC   treatment, insulin resistance, non-insulin-dependent diabetes mellitus
CC   (NIDDM), polycystic ovary syndrome, diseases of the gastrointestinal (GI)
CC   tract, inflammatory bowel disease, Crohn's disease, ulcerative colitis
CC   and bowel cancer. The nucleic acids are useful for studying the role of
CC   the PPAR-gamma gene in various diseases and disorders. The structure of
CC   PPAR-gamma enables genetic studies of PPAR- gamma mutations in humans,
CC   and evaluation of its role in disorders like insulin resistance, NIDDM,
CC   and diseases associated with altered adipose tissue function, like
CC   obesity and lipodystrophic syndromes. The nucleic acids are also useful
CC   for gene therapy and the production of transgenic animals, which are
CC   useful in screening assays. The control regions of the nucleic acids
CC   enable screening for modulators of the human PPAR-gamma gene, which are
CC   useful in designing drugs for treating disorders or diseases associated
CC   with the level of PPAR-gamma gene expression. The present sequence
CC   represents human PPAR-gamma-3 proximal promoter, exon A2 and intron A2
XX
SQ   Sequence 1463 BP; 461 A; 228 C; 284 G; 490 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 777;  DB 2;  Length 1463;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-183;
  Matches  777;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 427

Qy         61 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 487

Qy        121 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 547

Qy        181 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 607

Qy        241 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 667

Qy        301 GTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 GTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTC 727

Qy        361 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 787

Qy        421 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 847

Qy        481 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAAT 907

Qy        541 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 967

Qy        601 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 1027

Qy        661 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 1087

Qy        721 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 1144


RESULT 2
ADS36461
ID   ADS36461 standard; DNA; 158417 BP.
XX
AC   ADS36461;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Human autoimmune disease-related genomic DNA sequence - SEQ ID 1675.
XX
KW   single nucleotide polymorphism detection; SNP detection;
KW   rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW   systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW   thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW   glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW   primary systemic vasculitis; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004083403-A2.
XX
PD   30-SEP-2004.
XX
PF   18-MAR-2004; 2004WO-US008461.
XX
PR   18-MAR-2003; 2003US-0455444P.
PR   25-APR-2003; 2003US-0465241P.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Cargill M,  Begovich AB,  Alexander HC;
XX
DR   WPI; 2004-728480/71.
XX
PT   New isolated nucleic acid molecule comprises at least 8 contiguous
PT   nucleotides where one of the nucleotides is a single nucleotide
PT   polymorphism (SNP), useful for diagnosing or treating autoimmune
PT   diseases, e.g. rheumatoid arthritis.
XX
PS   Claim 16; SEQ ID NO 1675; 123pp; English.
XX
CC   The invention comprises amino acid and coding sequences containing
CC   genetic polymorphisms associated with an altered risk of developing an
CC   autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC   comprises a method of identifying an individual that has an altered risk
CC   of developing an autoimmune disease, comprising detecting a single
CC   nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC   and protein sequences of the invention are useful for diagnosing and
CC   treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC   diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC   bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC   anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC   myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC   present nucleic acid represents a human autoimmune disease-related
CC   genomic DNA sequence of the invention. NOTE: The present sequence is not
CC   shown in the specification, but has been retrieved from the WIPO website.
XX
SQ   Sequence 158417 BP; 47887 A; 30186 C; 31475 G; 48577 T; 0 U; 292 Other;

  Query Match             98.4%;  Score 764.2;  DB 13;  Length 158417;
  Best Local Similarity   98.6%;  Pred. No. 1.3e-179;
  Matches  766;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
```

```
Db      29693 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 29752

Qy         61 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 120

Db      29753 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTSATGGAAAATACAGCTATTCTAGG 29812

Qy        121 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 180

Db      29813 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 29872

Qy        181 TAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCA 240

Db      29873 TAGATGATCTGCTCTCTGGTTYGTCCTCTGAAAAGTCTGCTTCGYGAGGGGTGTGCTGCA 29932

Qy        241 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 300

Db      29933 TTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCAT 29992

Qy        301 GTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTC 360

Db      29993 GTGTCATCTAGAATGAAGTTATATTTTACTAAGGATCGTTTTTGCCATGTATAACTTCTC 30052

Qy        361 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 420

Db      30053 AAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCA 30112

Qy        421 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 480

Db      30113 GTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTA 30172

Qy        481 GTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAAT 540

Db      30173 GTGATGGAAATAGGAAAGTAGGTGAAGTGAWTCTAATAGATGTTTCTTTTATGAAATAAT 30232

Qy        541 TTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 600

Db      30233 TTTTAAAAGATTCTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTT 30292

Qy        601 ACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 660

Db      30293 ACTTCTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCC 30352

Qy        661 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 720

Db      30353 AATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGT 30412

Qy        721 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777

Db      30413 GTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 30469
```

```
RESULT 3
AAX19065
ID   AAX19065 standard; DNA; 468 BP.
XX
AC   AAX19065;
XX
DT   13-MAY-1999  (first entry)
XX
DE   Human PPAR-gamma-3 proximal promoter.
XX
KW   Human; peroxisome proliferator activated receptor gamma; PPAR-gamma;
KW   regulatory sequence; promoter; obesity; anorexia; lipoma; cachexia;
KW   lipodystrophy; liposarcoma; human immunodeficiency virus; HIV;
KW   insulin resistance; non-insulin-dependent diabetes mellitus;
KW   polycystic ovary syndrome; gastrointestinal tract; Crohn's disease;
KW   inflammatory bowel disease; ulcerative colitis; bowel cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9905161-A1.
XX
PD   04-FEB-1999.
XX
PF   24-JUL-1998;   98WO-US015411.
XX
PR   25-JUL-1997;   97US-0053692P.
XX
PA   (LIGA-) LIGAND PHARM INC.
PA   (INSP ) INST PASTEUR.
XX
PI   Briggs MR,  Saladin RS,  Auwerx J,  Fajas L;
XX
DR   WPI; 1999-142844/12.
XX
PT   Newly isolated nucleic acid comprising a control region of a human
PT   peroxisome proliferator activated receptor (PPAR) gamma gene - useful for
PT   identifying modulators that are useful in treating diseases associated
PT   with abnormal levels of human PPAR-gamma gene expression.
XX
PS   Disclosure; Page 88; 102pp; English.
XX
CC   The present invention describes an isolated, purified or enriched nucleic
CC   acid comprising a control region of a human peroxisome proliferator
CC   activated receptor gamma (PPAR-gamma) gene. The nucleic acids are useful
CC   for screening for agents capable of modulating the expression of a human
CC   PPAR-gamma gene. These agents (modulators) form pharmaceutical
CC   compositions that are useful for treating diseases associated with
CC   high/low levels of human PPAR-gamma gene expression. The diseases include
CC   obesity, anorexia, cachexia, lipodystrophy, lipomas, liposarcomas,
CC   abnormalities associated with anti-human immunodeficiency virus (HIV)
CC   treatment, insulin resistance, non-insulin-dependent diabetes mellitus
CC   (NIDDM), polycystic ovary syndrome, diseases of the gastrointestinal (GI)
CC   tract, inflammatory bowel disease, Crohn's disease, ulcerative colitis
CC   and bowel cancer. The nucleic acids are useful for studying the role of
CC   the PPAR-gamma gene in various diseases and disorders. The structure of
CC   PPAR-gamma enables genetic studies of PPAR- gamma mutations in humans,
CC   and evaluation of its role in disorders like insulin resistance, NIDDM,
CC   and diseases associated with altered adipose tissue function, like
CC   obesity and lipodystrophic syndromes. The nucleic acids are also useful
CC   for gene therapy and the production of transgenic animals, which are
CC   useful in screening assays. The control regions of the nucleic acids
CC   enable screening for modulators of the human PPAR-gamma gene, which are
CC   useful in designing drugs for treating disorders or diseases associated
CC   with the level of PPAR-gamma gene expression. The present sequence
CC   represents the human PPAR-gamma-3 proximal promoter
XX
SQ   Sequence 468 BP; 149 A; 63 C; 98 G; 158 T; 0 U; 0 Other;
```

```
  Query Match           49.3%;  Score 383;  DB 2;  Length 468;
  Best Local Similarity   99.7%;  Pred. No. 2.1e-85;
  Matches  394;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy        383 TAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCA 442

Db          1 TAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCA 60

Qy        443 TGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGG 502

Db         61 TGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGG 120

Qy        503 TGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCT 562

Db        121 TGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTT-AAAGATTGTCCAGCCCT 179

Qy        563 GCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCAT 622

Db        180 GCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCAT 239

Qy        623 TGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTT 682

Db        240 TGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTT 299

Qy        683 GGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACATGTGTG 742
```

```
Db        300 GGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACATGTGTG 359

Qy        743 TATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777
              |||||||||||||||||||||||||||||||||||
Db        360 TATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 394


RESULT 4
ADS39355
ID   ADS39355 standard; DNA; 201 BP.
XX
AC   ADS39355;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Human autoimmune disease-related SNP context sequence - SEQ ID 4569.
XX
KW   single nucleotide polymorphism detection; SNP detection;
KW   rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW   systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW   thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW   glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW   primary systemic vasculitis; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004083403-A2.
XX
PD   30-SEP-2004.
XX
PF   18-MAR-2004; 2004WO-US008461.
XX
PR   18-MAR-2003; 2003US-0455444P.
PR   25-APR-2003; 2003US-0465241P.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Cargill M,  Begovich AB,  Alexander HC;
XX
DR   WPI; 2004-728480/71.
XX
PT   New isolated nucleic acid molecule comprises at least 8 contiguous
PT   nucleotides where one of the nucleotides is a single nucleotide
PT   polymorphism (SNP), useful for diagnosing or treating autoimmune
PT   diseases, e.g. rheumatoid arthritis.
XX
PS   Claim 16; SEQ ID NO 4569; 123pp; English.
XX
CC   The invention comprises amino acid and coding sequences containing
CC   genetic polymorphisms associated with an altered risk of developing an
CC   autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC   comprises a method of identifying an individual that has an altered risk
CC   of developing an autoimmune disease, comprising detecting a single
CC   nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC   and protein sequences of the invention are useful for diagnosing and
CC   treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC   diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC   bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC   anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC   myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC   present DNA sequence represents a human autoimmune disease-related
CC   genomic-based SNP context sequence of the invention. NOTE: The present
CC   sequence is not shown in the specification, but has been retrieved from
CC   the WIPO website.
XX
SQ   Sequence 201 BP; 49 A; 43 C; 45 G; 63 T; 0 U; 1 Other;

  Query Match             25.8%;  Score 200.6;  DB 13;  Length 201;
  Best Local Similarity   99.5%;  Pred. No. 4.4e-40;
  Matches  200;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        102 AAAATACAGCTATTCTAGGATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAA 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db          1 AAAATACAGCTATTCTAGGATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAA 60

Qy        162 GACCCTAGGACCATTTACTTAGATGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCT 221
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db         61 GACCCTAGGACCATTTACTTAGATGATCTGCTCTCTGGTTYGTCCTCTGAAAAGTCTGCT 120

Qy        222 TCGTGAGGGGTGTGCTGCATTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCA 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCGTGAGGGGTGTGCTGCATTTGCCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCA 180

Qy        282 CCTTAGGCTTAATAACCATGT 302
              |||||||||||||||||||||
Db        181 CCTTAGGCTTAATAACCATGT 201


RESULT 5
ADS39165
ID   ADS39165 standard; DNA; 201 BP.
XX
AC   ADS39165;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Human autoimmune disease-related SNP context sequence - SEQ ID 4379.
XX
KW   single nucleotide polymorphism detection; SNP detection;
KW   rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW   systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW   thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW   glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW   primary systemic vasculitis; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004083403-A2.
XX
PD   30-SEP-2004.
XX
PF   18-MAR-2004; 2004WO-US008461.
XX
PR   18-MAR-2003; 2003US-0455444P.
PR   25-APR-2003; 2003US-0465241P.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Cargill M,  Begovich AB,  Alexander HC;
XX
DR   WPI; 2004-728480/71.
XX
PT   New isolated nucleic acid molecule comprises at least 8 contiguous
PT   nucleotides where one of the nucleotides is a single nucleotide
PT   polymorphism (SNP), useful for diagnosing or treating autoimmune
PT   diseases, e.g. rheumatoid arthritis.
XX
PS   Claim 16; SEQ ID NO 4379; 123pp; English.
XX
CC   The invention comprises amino acid and coding sequences containing
CC   genetic polymorphisms associated with an altered risk of developing an
CC   autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC   comprises a method of identifying an individual that has an altered risk
CC   of developing an autoimmune disease, comprising detecting a single
CC   nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC   and protein sequences of the invention are useful for diagnosing and
CC   treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC   diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC   bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC   anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC   myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC   present DNA sequence represents a human autoimmune disease-related
CC   genomic-based SNP context sequence of the invention. NOTE: The present
CC   sequence is not shown in the specification, but has been retrieved from
CC   the WIPO website.
XX
```

```
SQ   Sequence 201 BP; 48 A; 41 C; 46 G; 65 T; 0 U; 1 Other;

  Query Match             25.8%;  Score 200.6;  DB 13;  Length 201;
  Best Local Similarity   99.5%;  Pred. No. 4.4e-40;
  Matches  200;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        125 TTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACTTAGA 184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACTTAGA 60

Qy        185 TGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCATTTG 244
              ||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db         61 TGATCTGCTCTCTGGTTCGTCCTCTGAAAAGTCTGCTTCGYGAGGGGTGTGCTGCATTTG 120

Qy        245 CCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCATGTGT 304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTTGCCTAAGTGGTGTGGCACACAACTGTACTGTCACCTTAGGCTTAATAACCATGTGT 180

Qy        305 CATCTAGAATGAAGTTATATT 325
              |||||||||||||||||||||
Db        181 CATCTAGAATGAAGTTATATT 201


RESULT 6
ADS39169
ID   ADS39169 standard; DNA; 201 BP.
XX
AC   ADS39169;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Human autoimmune disease-related SNP context sequence - SEQ ID 4383.
XX
KW   single nucleotide polymorphism detection; SNP detection;
KW   rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW   systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW   thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW   glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW   primary systemic vasculitis; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004083403-A2.
XX
PD   30-SEP-2004.
XX
PF   18-MAR-2004; 2004WO-US008461.
XX
PR   18-MAR-2003; 2003US-0455444P.
PR   25-APR-2003; 2003US-0465241P.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Cargill M,  Begovich AB,  Alexander HC;
XX
DR   WPI; 2004-728480/71.
XX
PT   New isolated nucleic acid molecule comprises at least 8 contiguous
PT   nucleotides where one of the nucleotides is a single nucleotide
PT   polymorphism (SNP), useful for diagnosing or treating autoimmune
PT   diseases, e.g. rheumatoid arthritis.
XX
PS   Claim 16; SEQ ID NO 4383; 123pp; English.
XX
CC   The invention comprises amino acid and coding sequences containing
CC   genetic polymorphisms associated with an altered risk of developing an
CC   autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC   comprises a method of identifying an individual that has an altered risk
CC   of developing an autoimmune disease, comprising detecting a single
CC   nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC   and protein sequences of the invention are useful for diagnosing and
CC   treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC   diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC   bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC   anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC   myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC   present DNA sequence represents a human autoimmune disease-related
CC   genomic-based SNP context sequence of the invention. NOTE: The present
CC   sequence is not shown in the specification, but has been retrieved from
CC   the WIPO website.
XX
SQ   Sequence 201 BP; 50 A; 35 C; 47 G; 68 T; 0 U; 1 Other;

  Query Match             25.3%;  Score 196.6;  DB 13;  Length 201;
  Best Local Similarity   99.5%;  Pred. No. 4.3e-39;
  Matches  196;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 64

Qy         61 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 120
              ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db         65 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTSATGGAAAATACAGCTATTCTAGG 124

Qy        121 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 ATCCTTGAGCCTTTCATAAGAGATAAGGTTGTGAATCCTAAGACCCTAGGACCATTTACT 184

Qy        181 TAGATGATCTGCTCTCT 197
              |||||||||||||||||
Db        185 TAGATGATCTGCTCTCT 201


RESULT 7
ADS39163
ID   ADS39163 standard; DNA; 201 BP.
XX
AC   ADS39163;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Human autoimmune disease-related SNP context sequence - SEQ ID 4377.
XX
KW   single nucleotide polymorphism detection; SNP detection;
KW   rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW   systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW   thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW   glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW   primary systemic vasculitis; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004083403-A2.
XX
PD   30-SEP-2004.
XX
PF   18-MAR-2004; 2004WO-US008461.
XX
PR   18-MAR-2003; 2003US-0455444P.
PR   25-APR-2003; 2003US-0465241P.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Cargill M,  Begovich AB,  Alexander HC;
XX
DR   WPI; 2004-728480/71.
XX
PT   New isolated nucleic acid molecule comprises at least 8 contiguous
PT   nucleotides where one of the nucleotides is a single nucleotide
PT   polymorphism (SNP), useful for diagnosing or treating autoimmune
PT   diseases, e.g. rheumatoid arthritis.
XX
PS   Claim 16; SEQ ID NO 4377; 123pp; English.
XX
CC   The invention comprises amino acid and coding sequences containing
```

```
CC   genetic polymorphisms associated with an altered risk of developing an
CC   autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC   comprises a method of identifying an individual that has an altered risk
CC   of developing an autoimmune disease, comprising detecting a single
CC   nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC   and protein sequences of the invention are useful for diagnosing and
CC   treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC   diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC   bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC   anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC   myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC   present DNA sequence represents a human autoimmune disease-related
CC   genomic-based SNP context sequence of the invention. NOTE: The present
CC   sequence is not shown in the specification, but has been retrieved from
CC   the WIPO website.
XX
SQ   Sequence 201 BP; 62 A; 22 C; 46 G; 70 T; 0 U; 1 Other;

  Query Match             25.2%;  Score 195.8;  DB 13;  Length 201;
  Best Local Similarity   98.0%;  Pred. No. 6.8e-39;
  Matches  197;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        411 AGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTC 470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTC 60

Qy        471 ATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTT 530
              ||||||||||||||||||||||||||||||||||||:| ||||||||||||||||||
Db         61 ATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGAWTCTAATAGATGTTTCTTTT 120

Qy        531 ATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGG 590
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        121 ATGAAATAATTTTTAAAAGATTCTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGG 180

Qy        591 TCTGTTAGTTACTTTTAGAGA 611
              ||||||||||||| ||||||
Db        181 TCTGTTAGTTACTTCTAGAGA 201


RESULT 8
ABN80072
ID   ABN80072 standard; DNA; 5771 BP.
XX
AC   ABN80072;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human chemically modified disease associated gene SEQ ID NO 89.
XX
KW   Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW   heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW   dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW   antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200200927-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007536.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130908/17.
XX
PT   Novel nucleic acid useful for diagnosis and therapy of diseases
PT   associated with development genes such as diabetes, comprises a sequence
PT   of a segment of chemically pretreated DNA of genes associated with
PT   development.
XX
PS   Claim 1; SEQ ID NO 89; 27pp; English.
XX
CC   The invention relates to a nucleic acid (I) comprising a sequence at
CC   least 18 bases in length of a segment of chemically pretreated DNA (II)
CC   of genes associated with development selected from 87 genes listed in the
CC   specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
CC   sequences (ABN79984-ABN80333) or their complements. The invention is
CC   useful for the diagnosis or therapy of diseases associated with
CC   development genes, in particular disease related to homeobox containing
CC   genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC   associated with congenital heart disease, epilepsy, diseases related to
CC   histone deacetylation, Currarino syndrome, diseases related with the
CC   development of the brain and limb girdle muscular dystrophy and dwarfism.
CC   Oligomers specific to each of the genes are useful for detecting the
CC   methylation state of all CpG dinucleotides within the 350 sequences or
CC   (II) and their complementary sequences, as primer oligonucleotides for
CC   the amplification of the 350 sequences, (II) and/or their complements and
CC   as oligomer probes for detecting the cytosine methylation state and/or
CC   single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC   patent did not form part of the printed specification but is based on
CC   sequence information supplied to Derwent by the European Patent Office
XX
SQ   Sequence 5771 BP; 1877 A; 37 C; 980 G; 2877 T; 0 U; 0 Other;

  Query Match              6.4%;  Score 50;  DB 6;  Length 5771;
  Best Local Similarity   48.0%;  Pred. No. 0.033;
  Matches  143;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy        309 TAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTA 368
              || | |   ||| |||| | |||||||| | ||||   ||   ||   |||      ||
Db       1459 TAAATTTGTGTTTTATTGTGAAAAGGATAGGTTTTTTTATTATTATTATTTTTTTATTTT 1518

Qy        369 ACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGA 428
              ||   |  || || |  |||    |  || |||||| || |   |||  |||  |||
Db       1519 TTATTTTTGTATAGTATTTTTTTATTTGGTTAATTTTTATTTATGTTGTAAAGTTGTAGT 1578

Qy        429 GGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGA 488
              ||  |  |||  |||     ||        |  |   || |||   | || | ||||
Db       1579 TTTAGTGGAATTTTTGTAGGGAAAGTTTATATATTGTTATTTATTATAGATATTTATGGT 1638

Qy        489 AATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAA 548
              | || | |  |   |  || | ||  |  ||| |  || | |   | ||    | |
Db       1639 ATTATGTATTTTTTTTTTTGTAGTATTTGTTTATGATGGTTATTTTTTTTTATAAGAAGAG 1698

Qy        549 GATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTT 606
              | |||  ||          | ||     | ||  || | || |   ||| | |  |||
Db       1699 AAATGTAAAGGATAATTATAATTTAATGGTATTTTTATTTGTTAGATTAATAAAATTT 1756


RESULT 9
ABL32312
ID   ABL32312 standard; DNA; 18683 BP.
XX
AC   ABL32312;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 285.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 285; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 18683 BP; 5439 A; 130 C; 3594 G; 9520 T; 0 U; 0 Other;

  Query Match             6.3%;  Score 49.2;  DB 6;  Length 18683;
  Best Local Similarity   46.5%;  Pred. No. 0.072;
  Matches  228;  Conservative    0;  Mismatches  258;  Indels    4;  Gaps    2;

Qy        290 TTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATG 349
              |||| ||  |   ||  | |    |   |||| || ||     ||| ||| ||   ||
Db       6355 TTAAAAATAAGTCGTGGTTTGATAATTATAAATAATTAAATGTAGATAGTTATTATTATT 6414

Qy        350 TATAAATTTTCA--AACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTC 407
              |  || |||| |  ||   ||  |  ||  ||||   || ||   | | ||||
Db       6415 TTAAATTTTTGAGGAATTGTATTTTGTAATGTGTTTAAGTATTATATTATTTTGTTTGGA 6474

Qy        408 TTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGC 467
              | |   |      |  | ||  |  || |  |||  |     ||    |
Db       6475 ATTAAAATTGTGGTTTTGATAAGTTTTTTTTTTTAATTGATTTTTGTTATTTTTATTATT 6534

Qy        468 TTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCT 527
              |  |   | ||| | |  ||        ||  |    || | ||||  |   ||  || |
Db       6535 ATTGTAGTTTTTATTTAGTGATTATTATAATAATAAAAAATAATTTATAGTAATTATTTT 6594

Qy        528 TTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTG 587
              ||||  ||||| || ||| | ||| ||  |    || ||  ||| ||||  |  ||||
Db       6595 TTTAAAAAATATTTATTATATGATAGTTAATTTATGTATTTTTTTTGATTTAATATTTAA 6654

Qy        588 TGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGC--AAACA 645
              |  |   | |  || || |              |||||||    ||||  |     | | |
Db       6655 TATTAATTAATATAGTTGTGTTATTATTTTTATATTGTAGTTGAAGGGGATATTTAGAGA 6714

Qy        646 TTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCT 705
              |     | ||||    |  ||     || ||| ||  |    ||   ||| ||  |
Db       6715 ATGTTTTTAAAATATGTTTTAGGTTAAAAGTTTGTTGAGAGGAGTGGTTTTAGGTTGGTA 6774

Qy        706 AGGCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTT 765
              | | | |  | |||     | | | |   | |||  |||||   ||   | | | ||
Db       6775 TGATTTTTAAATTTTGTTATTAATTTAATTTTTTGTAAATATTTTTAAATATTTGTTTTG 6834

Qy        766 TTTATTGTTA 775
              ||| |  |||
Db       6835 TTTTTGATTA 6844
```

```
RESULT 10
ABL54333
ID   ABL54333 standard; DNA; 18683 BP.
XX
AC   ABL54333;
XX
DT   29-JUL-2002  (first entry)
XX
DE   Chemically treated apoptosis gene #17.
XX
KW   Apoptosis; HIV; Bloom syndrome; cardiopathy; neurodegenerative disorder;
KW   Herpes simplex virus; renal ischaemia; amyotrophic lateral sclerosis;
KW   cancer; ds.
XX
OS   Unidentified.
XX
PN   WO200177164-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP003969.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-017444/02.
XX
PT   Chemically modified sequences of genes associated with apoptosis are
PT   useful to determine methylation patterns of genomic DNA samples for
PT   diagnosis of associated diseases such as cancer.
XX
PS   Claim 1; Seq ID #33; 24pp; English.
XX
CC   This invention relates to chemically pre-treated DNA of genes associated
CC   with apoptosis. The nucleic acids are used to allocate patients for
CC   specific therapy for HIV infection, Bloom syndrome, cardiopathy, aging,
CC   neurodegenerative disorders, Herpes simplex virus infection, renal
CC   ischaemia, amyotrophic lateral sclerosis, solid tumours and cancers. This
CC   nucleotide sequence represents a chemically treated apoptosis gene. Even
CC   SEQ ID numbers are the complementary DNA strands to the odd SEQ ID
CC   numbers. The sequence data for this patent is not represented in the
CC   printed specification but is based on information supplied by the
CC   European patent office
XX
SQ   Sequence 18683 BP; 5439 A; 130 C; 3594 G; 9520 T; 0 U; 0 Other;

  Query Match             6.3%;  Score 49.2;  DB 6;  Length 18683;
  Best Local Similarity   46.5%;  Pred. No. 0.072;
  Matches  228;  Conservative    0;  Mismatches  258;  Indels    4;  Gaps    2;

Qy        290 TTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATG 349
              |||| ||  |   ||  | |    |   |||| || ||     ||| ||| ||   ||
Db       6355 TTAAAAATAAGTCGTGGTTTGATAATTATAAATAATTAAATGTAGATAGTTATTATTATT 6414

Qy        350 TATAAATTTTCA--AACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTC 407
              |  || |||| |  ||   ||  |  ||  ||||   || ||   | | ||||
Db       6415 TTAAATTTTTGAGGAATTGTATTTTGTAATGTGTTTAAGTATTATATTATTTTGTTTGGA 6474

Qy        408 TTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGC 467
              | |   |      |  | ||  |  || |  |||  |     ||    |
Db       6475 ATTAAAATTGTGGTTTTGATAAGTTTTTTTTTTTAATTGATTTTTGTTATTTTTATTATT 6534

Qy        468 TTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCT 527
              |  |   | ||| | |  ||        ||  |    || | ||||  |   ||  || |
Db       6535 ATTGTAGTTTTTATTTAGTGATTATTATAATAATAAAAAATAATTTATAGTAATTATTTT 6594
```

```
Qy       528 TTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTG 587
Db      6595 TTTAAAAAATATTTATTATATGATAGTTAATTTATGTATTTTTTTTGATTTAATATTTAA 6654

Qy       588 TGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGC--AAACA 645
Db      6655 TATTAATTAATATAGTTGTGTTATTATTTTTATATTGTAGTTGAAGGGGATATTTAGAGA 6714

Qy       646 TTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCT 705
Db      6715 ATGTTTTTAAAATATGTTTTAGGTTAAAAGTTTGTTGAGAGGAGTGGTTTTAGGTTGGTA 6774

Qy       706 AGGCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTT 765
Db      6775 TGATTTTTAAATTTTGTTATTAATTTAATTTTTTGTAAATATTTTTAAATATTTGTTTTG 6834

Qy       766 TTTATTGTTA 775
Db      6835 TTTTTGATTA 6844
```

RESULT 11
ABK39937
ID ABK39937 standard; DNA; 11422 BP.
XX
AC ABK39937;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #9 strand 2.
XX
KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 18; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 U; 0 Other;

Query Match 6.3%; Score 49; DB 6; Length 11422;
Best Local Similarity 46.4%; Pred. No. 0.071;
Matches 235; Conservative 0; Mismatches 265; Indels 7; Gaps 2;

```
Qy       272 TGTACTGTCACCTTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAA 331
Db      6619 TTTTTTTTAATTTTAAATTTTAAAAAAATTTTTTTTTTTTTAAGAATATATATATTTTTT 6678

Qy       332 AGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTT 391
Db      6679 ATGTTTGTTTTAAGATTTAAAGACGTATATATTATTTTAAAGTAGTGTTATTA-----TT 6733

Qy       392 TAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATA 451
Db      6734 GAAATTTATTATAGTTTTAATTTGTGTTAATTATATAGTTAGAGAAGATAGTAATATAGA 6793

Qy       452 AAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGAT 511
Db      6794 AAATAAGAAATATAATGTTTTAGATATAAGTTATGTAAATTAAAAATTTAGTTTTATTAA 6853

Qy       512 TTTAATAGATGTTTCTTTTATGAAATAATTTTTAA--AAGATTGTCCAGCCCTGCATGAT 569
Db      6854 TTGAATGGATGTTTTTGATATGTATTTTTTTTTAATTTTTATATTTTAATATTTATTTAA 6913

Qy       570 TTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGC 629
Db      6914 GTTATATAAATTTTAGTATTGGTAATTTTTTATTTAGAAGATTTTGTAGGTAGTTATATA 6973

Qy       630 TCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATG 689
Db      6974 TGATGAGAAGTTATAATTTTTTGGTTAAGATAGGAAAAATAAAATGAATAATTTTTATTA 7033

Qy       690 TACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTG 749
Db      7034 AATTTTATAAAAGAGATATATTTGTTAATTTGTAGTGTATATGTTTTGGAGAATATATTT 7093

Qy       750 AAAATCTTGTCTCTTTTTTATTGTTAA 776
Db      7094 ATGTTTTTGGCGTTTATTTTTTTTGAA 7120
```

RESULT 12
ABL32219
ID ABL32219 standard; DNA; 11422 BP.
XX
AC ABL32219;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 192.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.

XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 192; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 U; 0 Other;

Query Match 6.3%; Score 49; DB 6; Length 11422;
Best Local Similarity 46.4%; Pred. No. 0.071;
Matches 235; Conservative 0; Mismatches 265; Indels 7; Gaps 2;

```
Qy       272 TGTACTGTCACCTTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAA 331
Db      6619 TTTTTTTAATTTTAAATTTTAAAAAAATTTTTTTTTTTTTAAGAATATATATATTTTTT 6678

Qy       332 AGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTT 391
Db      6679 ATGTTTGTTTTAAGATTTAAAGACGTATATATTATTTTAAAGTAGTGTTATTA-----TT 6733

Qy       392 TAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATA 451
Db      6734 GAAATTTATTATAGTTTTAATTTGTGTTAATTATATAGTTAGAGAAGATAGTAATATAGA 6793

Qy       452 AAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGAT 511
Db      6794 AAATAAGAAATATAATGTTTTAGATATAAGTTATGTAAATTAAAAATTTAGTTTTATTAA 6853

Qy       512 TTTAATAGATGTTTCTTTTATGAAATAATTTTTAA--AAGATTGTCCAGCCCTGCATGAT 569
Db      6854 TTGAATGGATGTTTTTGATATGTATTTTTTTTTAATTTTTATATTTTAATATTTATTTAA 6913

Qy       570 TTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGC 629
Db      6914 GTTATATAAATTTTAGTATTGGTAATTTTTTATTTAGAAGATTTTGTAGGTAGTTATATA 6973

Qy       630 TCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATG 689
Db      6974 TGATGAGAAGTTATAATTTTTTGGTTAAGATAGGAAAAATAAAATGAATAATTTTTATTA 7033

Qy       690 TACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTG 749
Db      7034 AATTTTATAAAAGAGATATATTTGTTAATTTGTAGTGTATATGTTTTGGAGAATATATTT 7093

Qy       750 AAAATCTTGTCTCTTTTTTATTGTTAA 776
Db      7094 ATGTTTTTGGCGTTTATTTTTTTTGAA 7120
```

RESULT 13
ABK31423
ID ABK31423 standard; DNA; 13123 BP.
XX
AC ABK31423;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified complementary DNA #133.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007472.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX
PS Claim 1; SEQ ID NO 266; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are
CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
XX
SQ Sequence 13123 BP; 3999 A; 255 C; 2547 G; 6320 T; 0 U; 2 Other;

Query Match 6.2%; Score 48.4; DB 6; Length 13123;
Best Local Similarity 49.1%; Pred. No. 0.1;
Matches 157; Conservative 0; Mismatches 161; Indels 2; Gaps 1;

```
Qy       284 TTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTT 343
Db      2473 TTATGTATAAAATTTTTATTTTAATTTAGTTTAAATTTTATTTTTAAAAGTATTTTTGTT 2532

Qy       344 GCCATGTATAAATTTTCAAACATTAACTT--TCAGGGTTATTAATCCTTTTAAGGTCTAG 401
Db      2533 TTTTGTTTTAATTATTTATGTATTTTTTTAGTTTTGATAGCGTTTAGTTTTATTGATTTT 2592
```

```
Qy        402 TTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAA 461
              ||||| ||    |       | |||||  |||||     ||  | | | | |    ||  |
Db       2593 TTTTTATTGTTTTGAAAGATTAATATGGGTATTAATTTTTTTATAAGTTTATTTATTTAT 2652

Qy        462 AGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGAT 521
                |  ||| || |||    | |   ||||  |  ||       |  ||| | |  ||
Db       2653 TAAAAATATATTTATATTATAAGTTGTTTTAGGTATTTATTAAGTATTTTTTAAGTTTAT 2712

Qy        522 GTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATC 581
              ||| ||||||  | |   ||| ||  |||||          ||  ||| | ||   |
Db       2713 TTTTTTTTTATTTATTGTATTTGAATTAATTGTTAGAGTTAAGATATTTTTTTATTTTTT 2772

Qy        582 ATTTTGTGGTCTGTTAGTTA 601
              |||| ||||  |||   ||
Db       2773 TTTTTTTGGTTAGTTTTGTA 2792
```

RESULT 14
ABL54364
ID ABL54364 standard; DNA; 13123 BP.
XX
AC ABL54364;
XX
DT 29-JUL-2002 (first entry)
XX
DE Chemically treated apoptosis gene complementary to gene #32.
XX
KW Apoptosis; HIV; Bloom syndrome; cardiopathy; neurodegenerative disorder;
KW Herpes simplex virus; renal ischaemia; amyotrophic lateral sclerosis;
KW cancer; ds.
XX
OS Unidentified.
XX
PN WO200177164-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003969.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017444/02.
XX
PT Chemically modified sequences of genes associated with apoptosis are
PT useful to determine methylation patterns of genomic DNA samples for
PT diagnosis of associated diseases such as cancer.
XX
PS Claim 1; Seq ID #64; 24pp; English.
XX
CC This invention relates to chemically pre-treated DNA of genes associated
CC with apoptosis. The nucleic acids are used to allocate patients for
CC specific therapy for HIV infection, Bloom syndrome, cardiopathy, aging,
CC neurodegenerative disorders, Herpes simplex virus infection, renal
CC ischaemia, amyotrophic lateral sclerosis, solid tumours and cancers. This
CC nucleotide sequence represents a chemically treated apoptosis gene. Even
CC SEQ ID numbers are the complementary DNA strands to the odd SEQ ID
CC numbers. The sequence data for this patent is not represented in the
CC printed specification but is based on information supplied by the
CC European patent office
XX
SQ Sequence 13123 BP; 3999 A; 255 C; 2547 G; 6320 T; 0 U; 2 Other;

Query Match 6.2%; Score 48.4; DB 6; Length 13123;
Best Local Similarity 49.1%; Pred. No. 0.1;
Matches 157; Conservative 0; Mismatches 161; Indels 2; Gaps 1;

```
Qy        284 TTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTT 343
              ||| |  ||| |    | | | |  |     | || || |||||| ||||| ||  || ||
Db       2473 TTATGTATAAAATTTTTATTTTAATTTAGTTTAAATTTTATTTTTAAAAGTATTTTTGTT 2532

Qy        344 GCCATGTATAAATTTTCAAACATTAACTT--TCAGGGTTATTAATCCTTTTAAGGTCTAG 401
                  | ||| | || |  |||   ||  |  | |      |  |||||  |  |
Db       2533 TTTTGTTTTAATTATTTATGTATTTTTTTAGTTTTGATAGCGTTTAGTTTTATTGATTTT 2592

Qy        402 TTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAA 461
              ||||| ||    |       | |||||  |||||     ||  | | | | |    ||  |
Db       2593 TTTTTATTGTTTTGAAAGATTAATATGGGTATTAATTTTTTTATAAGTTTATTTATTTAT 2652

Qy        462 AGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGAT 521
                |  ||| || |||    | |   ||||  |  ||       |  ||| | |  ||
Db       2653 TAAAAATATATTTATATTATAAGTTGTTTTAGGTATTTATTAAGTATTTTTTAAGTTTAT 2712

Qy        522 GTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATC 581
              ||| ||||||  | |   ||| ||  |||||          ||  ||| | ||   |
Db       2713 TTTTTTTTTATTTATTGTATTTGAATTAATTGTTAGAGTTAAGATATTTTTTTATTTTTT 2772

Qy        582 ATTTTGTGGTCTGTTAGTTA 601
              |||| ||||  |||   ||
Db       2773 TTTTTTTGGTTAGTTTTGTA 2792
```

RESULT 15
ABK31322
ID ABK31322 standard; DNA; 6880 BP.
XX
AC ABK31322;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified DNA #83.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007472.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX
PS Claim 1; SEQ ID NO 165; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,

CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are
CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
XX
SQ Sequence 6880 BP; 2032 A; 49 C; 1380 G; 3419 T; 0 U; 0 Other;

Query Match 6.0%; Score 47; DB 6; Length 6880;
Best Local Similarity 51.8%; Pred. No. 0.19;
Matches 131; Conservative 0; Mismatches 120; Indels 2; Gaps 1;

```
Qy        301 GTGTCATCTAGAATGAAGTTATATTTTAAAA--AGGATCGTTTTTGCCATGTATAAATTT 358
              |||| || | |  || || ||  | ||||||   ||    |||||   | ||||||||
Db        722 GTGTTATTTTGTGTGTAGATAGTTGTTAAAATTTGGTGTTTTTTTTTAGAGGTATAAATGA 781

Qy        359 TCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTG 418
              |  |   ||   |||     || ||  |  ||||    | | ||||||    ||    |
Db        782 TGTAGGTTTTTATTTTGTTATTTTTTTTTGTTTTGTTTTTTTGTTTTTTGAGAGGTATTA 841

Qy        419 CAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGT 478
               | |||| ||  | |  |   | || |   || |    |   || |   ||  ||  |||
Db        842 TATTAATGGAAATTTAATATATAATTTATAATTATGTTTTTGAAAGTTTTTTTTTTTTTGT 901

Qy        479 TAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATA 538
              |  | ||    | |    ||||| |  || | | | | | | | ||  || |||| |  |
Db        902 TTTTAATATTTAAATTTTAGTAGTTTTAGGGGTATAAGTGGTTTTTGATTATATGGATGA 961

Qy        539 ATTTTTAAAAGAT 551
              ||| |  | | ||
Db        962 ATTGTATAGAAAT 974
```

Search completed: November 9, 2005, 00:12:41
Job time : 631.342 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 15:16:42 ; Search time 197.983 Seconds
(without alignments)
6421.701 Million cell updates/sec

Title: US-09-463-542-34_COPY_368_1144
Perfect score: 777
Sequence: 1 tcatgtaggtaagactgtgt.....gtctctttttattgttaag 777

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	45.4	5.8	1743	4	US-09-248-796A-11015	Sequence 11015, A
	2	44.4	5.7	54576	4	US-09-949-016-15954	Sequence 15954, A
	3	44.4	5.7	54576	4	US-09-949-016-15955	Sequence 15955, A
	4	41.8	5.4	231129	4	US-09-949-016-16110	Sequence 16110, A
	5	41.8	5.4	266293	4	US-09-949-016-11934	Sequence 11934, A
c	6	41.4	5.3	601	4	US-09-949-016-156535	Sequence 156535,
c	7	40.4	5.2	1141	4	US-09-806-708B-22	Sequence 22, Appl
	8	40.4	5.2	360470	4	US-09-949-016-13173	Sequence 13173, A
	9	39.8	5.1	175236	4	US-09-949-016-14353	Sequence 14353, A
c	10	39.8	5.1	298336	4	US-09-949-016-16600	Sequence 16600, A
	11	39.6	5.1	1141	4	US-09-806-708B-22	Sequence 22, Appl
c	12	39	5.0	601	4	US-09-949-016-142482	Sequence 142482,
	13	39	5.0	181429	4	US-09-949-016-12372	Sequence 12372, A
	14	39	5.0	181430	4	US-09-949-016-15772	Sequence 15772, A
c	15	38.4	4.9	1783	4	US-09-679-409-9	Sequence 9, Appli
c	16	38.4	4.9	1898	4	US-09-679-409-8	Sequence 8, Appli
c	17	38.4	4.9	1923	4	US-09-679-409-12	Sequence 12, Appl
c	18	38.4	4.9	1981	4	US-09-679-409-6	Sequence 6, Appli
c	19	38.4	4.9	1992	4	US-09-679-409-7	Sequence 7, Appli
c	20	38.4	4.9	2075	4	US-09-679-409-11	Sequence 11, Appl
c	21	38.4	4.9	2132	4	US-09-679-409-5	Sequence 5, Appli
c	22	38.4	4.9	2215	4	US-09-679-409-4	Sequence 4, Appli
c	23	38.4	4.9	2226	4	US-09-679-409-2	Sequence 2, Appli
c	24	38.4	4.9	2308	4	US-09-679-409-13	Sequence 13, Appl
c	25	38.4	4.9	2309	4	US-09-679-409-10	Sequence 10, Appl
c	26	38.4	4.9	2431	4	US-09-679-409-3	Sequence 3, Appli
c	27	38.4	4.9	2612	4	US-09-679-409-15	Sequence 15, Appl
c	28	38.4	4.9	2706	4	US-09-679-409-14	Sequence 14, Appl
c	29	38.4	4.9	319608	4	US-09-539-333D-1	Sequence 1, Appli
c	30	38.4	4.9	319608	4	US-09-679-409-1	Sequence 1, Appli
	31	38.2	4.9	601	4	US-09-949-016-18854	Sequence 18854, A
	32	38.2	4.9	601	4	US-09-949-016-56719	Sequence 56719, A
	33	38.2	4.9	832	4	US-09-621-976-2813	Sequence 2813, Ap
c	34	38.2	4.9	4291	2	US-08-417-210A-80	Sequence 80, Appl
c	35	38.2	4.9	4291	4	US-09-136-159A-80	Sequence 80, Appl
c	36	38.2	4.9	4857	2	US-08-566-398-16	Sequence 16, Appl
c	37	38.2	4.9	6628	3	US-08-815-809-3	Sequence 3, Appli
c	38	38.2	4.9	6649	2	US-08-816-155B-5	Sequence 5, Appli
c	39	38.2	4.9	6649	3	US-09-079-587-5	Sequence 5, Appli
c	40	38.2	4.9	7091	2	US-08-658-665-40	Sequence 40, Appl
c	41	38.2	4.9	7091	3	US-08-796-101-4	Sequence 4, Appli
c	42	38.2	4.9	7091	3	US-09-085-273-40	Sequence 40, Appl
c	43	38.2	4.9	7091	4	US-09-916-963-40	Sequence 40, Appl
c	44	38.2	4.9	7351	1	US-08-224-391-83	Sequence 83, Appl
c	45	38.2	4.9	7351	1	US-08-484-304-83	Sequence 83, Appl

ALIGNMENTS

RESULT 1
US-09-248-796A-11015/c
; Sequence 11015, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 11015
; LENGTH: 1743
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-11015

Query Match 5.8%; Score 45.4; DB 4; Length 1743;
Best Local Similarity 48.6%; Pred. No. 0.028;
Matches 124; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
Qy        492 AGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGAT 551
              ||  || ||  | |  |  | | ||||  ||  | ||| || ||      || |||| ||
Db       1431 AGTTAAATATTTAATTTTTTGTAAATAATTGAATTTTTCATTAATATTAATTAAAAAAAT 1372

Qy        552 TGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGA 611
              | | |    ||  | | | ||  |   | | ||  ||||   | | | | |||  || |
Db       1371 TTTGAAAATATGATTTAATAATTTTCGTTTAATTAATGGTGGATAATTGATTTTATGATA 1312

Qy        612 ATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGG 671
              | || ||||  |  | | |  |  | ||  |||||||| ||      |  |||
Db       1311 TAATAATGCATAATTTGGTAATATATTGAAATTTTTCAGAATTAATCGTTCAACGGAAAC 1252

Qy        672 TAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCT 731
              |||||   |||  |  |  ||||  |  | |     |  | ||  ||   || |
Db       1251 TAAATAACCTTGATAAATTTTTTTTTTAAATACAATTTCATAAAAATTAGTCTTAATCCG 1192

Qy        732 GAACATGTGTGTATA 746
              |  || ||  | |
Db       1191 TGAGGTGGGTTGAGA 1177
```

RESULT 2
US-09-949-016-15954
; Sequence 15954, Application US/09949016

```
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15954
;   LENGTH: 54576
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(54576)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15954

  Query Match             5.7%;  Score 44.4;  DB 4;  Length 54576;
  Best Local Similarity   50.0%;  Pred. No. 0.16;
  Matches  137;  Conservative    0;  Mismatches  136;  Indels    1;  Gaps    1;

Qy        271 CTGTACTGTCACCTTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAA 330
              ||||| ||||  |||||  | | |||       | |||      |||||    |||   ||
Db      29158 CTGTAGTGTCTACTTAGTTTGACTAAGTGGTAGACATTGTACATGAAAAACTATGGAAAC 29217

Qy        331 AAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTT 390
              |    |  ||     |||    | || | |||  ||    |    |  ||||||| ||   |||
Db      29218 TAATGTTATTGCGTCCAGAAAGAATTGTTCTTAC-GGATGGTATAGGGTTAGTACATCTT 29276

Qy        391 TTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACAT 450
              ||| | ||  | | | |   || ||  |        || |   |  |  |   | |   |
Db      29277 GTAAAGGCTGTTCTATTTCTGGTTTGCCCTTATCATGAAGGCACAGCCCTTCAGGGGTCT 29336

Qy        451 AAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGA 510
              || |         |   ||||||| | ||  |  | | | ||  ||| ||   ||  ||
Db      29337 CAATGGAAAGCCTGAAATTCATTCTTTTTTCTCCTAGCACTATCAAACTACTGGATTTGT 29396

Qy        511 TTTTAATAGATGTTTCTTTTATGAAATAATTTTT 544
                |||||   |    |||||| ||| || ||| |
Db      29397 AGTTAATTTTTTAGCCTTTTAGGAACTATTTTCT 29430


RESULT 3
US-09-949-016-15955
; Sequence 15955, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15955
;   LENGTH: 54576
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(54576)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15955

  Query Match             5.7%;  Score 44.4;  DB 4;  Length 54576;
  Best Local Similarity   50.0%;  Pred. No. 0.16;
  Matches  137;  Conservative    0;  Mismatches  136;  Indels    1;  Gaps    1;

Qy        271 CTGTACTGTCACCTTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAA 330
              ||||| ||||  |||||  | | |||       | |||      |||||    |||   ||
Db      29158 CTGTAGTGTCTACTTAGTTTGACTAAGTGGTAGACATTGTACATGAAAAACTATGGAAAC 29217

Qy        331 AAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTT 390
              |    |  ||     |||    | || | |||  ||    |    |  ||||||| ||   |||
Db      29218 TAATGTTATTGCGTCCAGAAAGAATTGTTCTTAC-GGATGGTATAGGGTTAGTACATCTT 29276

Qy        391 TTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACAT 450
              ||| | ||  | | | |   || ||  |        || |   |  |  |   | |   |
Db      29277 GTAAAGGCTGTTCTATTTCTGGTTTGCCCTTATCATGAAGGCACAGCCCTTCAGGGGTCT 29336

Qy        451 AAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGA 510
              || |         |   ||||||| | ||  |  | | | ||  ||| ||   ||  ||
Db      29337 CAATGGAAAGCCTGAAATTCATTCTTTTTTCTCCTAGCACTATCAAACTACTGGATTTGT 29396

Qy        511 TTTTAATAGATGTTTCTTTTATGAAATAATTTTT 544
                |||||   |    |||||| ||| || ||| |
Db      29397 AGTTAATTTTTTAGCCTTTTAGGAACTATTTTCT 29430


RESULT 4
US-09-949-016-16110
; Sequence 16110, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16110
;   LENGTH: 231129
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16110

  Query Match             5.4%;  Score 41.8;  DB 4;  Length 231129;
  Best Local Similarity   45.5%;  Pred. No. 1.3;
  Matches  148;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        285 TAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTG 344
              ||   || ||    || |   || || | |  | |||||| | | | |   |  | |||
Db      72576 TATATTTTATTTATATATAAAATATATATTTTATTTATATATAATATATATTTATATTTT 72635

Qy        345 CCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTT 404
              || |||| ||  |  |   ||| | | |     |  ||    |||  |   |    | |
Db      72636 TTATTTATATATAATATATATTTATATATTATATATTATATATTTATATATTATATAT 72695

Qy        405 TTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGG 464
              |   ||  | | |  | ||      |||    ||| ||||   ||| |  ||  | |
Db      72696 TATATATATTTATATATTATATATAATATATATATTTATGTATTATATATTATATATATT 72755
```

```
Qy        465 GGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTT 524
              |   | || |   |    | | ||  |  ||  |    |   | | ||| || ||
Db      72756 TATATATGTAATATATATTTATGTATTATATATTTATATATTATATATATTATATATATT 72815

Qy        525 TCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATT 584
              |   | | ||||| | ||| | ||  |  |    |     || ||| ||  ||  ||
Db      72816 ATATACTTAATATAATATATAATATATATTATATATATTGTATATATATAATATATATTT 72875

Qy        585 TTGTGGTCTGTTAGTTACTTTTAGA 609
              | |  | | |   | |  | || |
Db      72876 ATATATTTTATATATAATATATATA 72900


RESULT 5
US-09-949-016-11934
; Sequence 11934, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11934
;   LENGTH: 266293
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-11934

  Query Match             5.4%;  Score 41.8;  DB 4;  Length 266293;
  Best Local Similarity   45.5%;  Pred. No. 1.3;
  Matches  148;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        285 TAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTG 344
              ||   || ||   || |  || || | |  | |||||| | | |  |   | | |||
Db      59740 TATATTTTATTTATATATAAAATATATATTTTATTTATATATAATATATATTTATATTTT 59799

Qy        345 CCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTT 404
              || |||| ||  |  |   |||  | | |    |  ||    |||  |  |    | |
Db      59800 TTATTTATATATAATATATATTTATATATTATATATTATATATATTTATATATTATATAT 59859

Qy        405 TTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGG 464
              |   ||  | | |  | ||        |||    |||  ||||   |||  |  ||  | |
Db      59860 TATATATATTTATATATTATATATAATATATATATTTATGTATTATATATTATATATATT 59919

Qy        465 GGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTT 524
              |   | || |   |    | | ||  |  ||  |    |   | | ||| || ||
Db      59920 TATATATGTAATATATATTTATGTATTATATATTTATATATTATATATATTATATATATT 59979

Qy        525 TCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATT 584
              |   | | ||||| | ||| | ||  |  |    |     || ||| ||  ||  ||
Db      59980 ATATACTTAATATAATATATAATATATATTATATATATTGTATATATATAATATATATTT 60039

Qy        585 TTGTGGTCTGTTAGTTACTTTTAGA 609
              | |  | | |   | |  | || |
Db      60040 ATATATTTTATATATAATATATATA 60064


RESULT 6
US-09-949-016-156535/c
; Sequence 156535, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 156535
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-156535

  Query Match             5.3%;  Score 41.4;  DB 4;  Length 601;
  Best Local Similarity   45.2%;  Pred. No. 0.23;
  Matches  147;  Conservative    1;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        285 TAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTG 344
              ||   || ||     || |  || || | |  | |||||| | | |  |   | | |||
Db      438 TATATTTTATTTATATATAAAATATATATTTTATTTATATATAATATATATTTATATTTT 379

Qy        345 CCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTT 404
              || |||| ||  |  |   |||  | | |    |  ||    |||  |  |    | |
Db      378 TTATTTATATATAATATATATTTATATATTATATATTATATATATTTATATATTATATAT 319

Qy        405 TTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGG 464
              |   ||  | | |  | :|        |||    |||  ||||   |||  |  ||  | |
Db      318 TATATATATTTATATATWATATATAATATATATATTTATGTATTATATATTATATATATT 259

Qy        465 GGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTT 524
              |   | || |   |    | | ||  |  ||  |    |   | | ||| || ||
Db      258 TATATATGTAATATATATTTATGTATTATATATTTATATATTATATATATTATATATATT 199

Qy        525 TCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATT 584
              |   | | ||||| | ||| | ||  |  |    |     || ||| ||  ||  ||
Db      198 ATATACTTAATATAATATATAATATATATTATATATATTGTATATATATAATATATATTT 139

Qy        585 TTGTGGTCTGTTAGTTACTTTTAGA 609
              | |  | | |   | |  | || |
Db      138 ATATATTTTATATATAATATATATA 114


RESULT 7
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             5.2%;  Score 40.4;  DB 4;  Length 1141;
```

Best Local Similarity 9.3%; Pred. No. 0.52;
Matches 34; Conservative 155; Mismatches 178; Indels 0; Gaps 0;

```
Qy        316 AAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCA 375
Db        695 WWNNTDARRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNNNNNNNNNSCC 636

Qy        376 GGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCG 435
Db        635 TCTRMMTMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMRCRDVTYTRNNT 576

Qy        436 TCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGA 495
Db        575 YCKSYAHSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRWAGTMWWRHWNN 516

Qy        496 AAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTC 555
Db        515 NNTDTRYYWWWKRWARBTTTVYDSMCNAKSMWRGNNWRAMKMWWAANNDAGAMDHWTYWM 456

Qy        556 CAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAG 615
Db        455 GNNTMWMRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWTWKYMWKAACNNNNBKA 396

Qy        616 AAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAA 675
Db        395 MYMRVAWMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNNNNNNWRBCKTTSWMWW 336

Qy        676 TTTATTT 682
Db        335 MDHMNTH 329
```

RESULT 8
US-09-949-016-13173
; Sequence 13173, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13173
; LENGTH: 360470
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13173

Query Match 5.2%; Score 40.4; DB 4; Length 360470;
Best Local Similarity 46.5%; Pred. No. 3.5;
Matches 131; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

```
Qy        292 AATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTA 351
Db     176767 AAGATGCACCTATAAGATACAAAACTTTAATTAATACAAAACGTTAGTAGTGTGTATTTT 176826

Qy        352 TAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAA 411
Db     176827 CACTTCATCCATTTACTTTGCTCAGGTTGATTACATAGTTTGGGGTATATACTATGTGAA 176886

Qy        412 GTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCA 471
Db     176887 GATACACCTATAAGACACAAAACTTTAATTCAAGAAAATGTTAGACATCTAACATTTATG 176946

Qy        472 TTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTA 531
Db     176947 GTTATATAATTTATTGTCATGTAATTTATTGTTATGTAATTTATTGTTATTTTTCTCTTA 177006

Qy        532 TGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTAT 573
Db     177007 TTTAGTCATTTGATAAATATTGAGCTACCTCCTATGCCTTAT 177048
```

RESULT 9
US-09-949-016-14353
; Sequence 14353, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14353
; LENGTH: 175236
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14353

Query Match 5.1%; Score 39.8; DB 4; Length 175236;
Best Local Similarity 48.1%; Pred. No. 4;
Matches 113; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

```
Qy        146 AGGTTGTGAATCCTAAGACCCTAGGACCATTTACTTAGATGATCTGCTCTCTGGTTCGTC 205
Db     107752 AGCTCTTGTCTGATATTACTGCTGCAGGCTTTATTTTGATTAGCATTTTCCTAGTATGTA 107811

Qy        206 CTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCATTTGCCTTGCCTAAGTGGTGTGGCA 265
Db     107812 CATTTTGTCTTTTCAGTCTTTCTATGTGTTGTGTTTTAGTTGTGACTCTTAGAAGATGTA 107871

Qy        266 CACAACTGTACTGTCACCTTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATT 325
Db     107872 TATAGCTAGATTGTTAAAATTCACTCTATTATTATTTAGTATGTTGAATTTAATCCAATT 107931

Qy        326 TTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTT 380
Db     107932 ATATAGATGATCATTACTGATCTATTTAGACTTATTACTATCTGATTTTGTGGTT 107986
```

RESULT 10
US-09-949-016-16600/c
; Sequence 16600, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16600
;   LENGTH: 298336
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(298336)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16600

  Query Match             5.1%;  Score 39.8;  DB 4;  Length 298336;
  Best Local Similarity   47.7%;  Pred. No. 4.7;
  Matches  116;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy        528 TTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTG 587
              | |||   |||          | ||| | |  ||   |  || || ||     |  | | |
Db     273892 TATATATTATATACACGTATAGAATATATAGTTATATATTATATACACGTATAGAATATA 273833

Qy        588 TGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATT 647
              | ||  ||||  |||   || |||||| | ||   | ||  |  |  |||   | | |
Db     273832 TAGTTAGTTATATACACGTATAGAATATATAGTTATATACACGTATAGAATATATAGTTA 273773

Qy        648 CAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAG 707
              | | |  |  |   || ||||  || | |||| |           | | | |
Db     273772 TATATTATATACACGTATAGAATATATAGTTATATATTATATACACGTATAGAATATATA 273713

Qy        708 GCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTT 767
              | | | |||| || || |   | |    || |     ||| ||| || | | || | | ||
Db     273712 GTTATATACACGTATAGAATATATAGTTATATACACGTATAGAATATATAGTATATATTA 273653

Qy        768 TAT 770
              |||
Db     273652 TAT 273650


RESULT 11
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE:  2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 1141
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1)..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

  Query Match             5.1%;  Score 39.6;  DB 4;  Length 1141;
  Best Local Similarity   8.8%;  Pred. No. 0.85;
  Matches   55;  Conservative  252;  Mismatches  320;  Indels    0;  Gaps    0;

Qy        145 AAGGTTGTGAATCCTAAGACCCTAGGACCATTTACTTAGATGATCTGCTCTCTGGTTCGT 204
              :: : : :: :::|     : ::|:::  ::|  : |:  :  :    : :   :  ::|
Db         77 RWWKGWYKKKWYBCANNTSBRYHARRWKDMKTAYBMTMTNKWGKTGWRHRYWRWRAMBDT 136

Qy        205 CCTCTGAAAAGTCTGCTTCGTGAGGGGTGTGCTGCATTTGCCTTGCCTAAGTGGTGTGGC 264
              : :::: |:    | :  ::::::: :: ::    || ::  |: :::     :|:|  :
Db        137 VDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDDDTKYHMWNNNGCBTVTWMV 196

Qy        265 ACACAACTGTACTGTCACCTTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATAT 324
              :: :: :::    : :: ::  :::   :::: ::  :  :  ::: : ::: : : ::
Db        197 RYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRGRMRNYMVAWB 256

Qy        325 TTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTA 384
              | ::::   |:|:  :    :       | |  :   : :  | :  ||   : : |::
Db        257 TAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAKYWGWNRABVNSTCTTWKSKTTKV 316

Qy        385 ATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATG 444
              :|:|:    : |:   ::::::   |:| ::       : : :|: :|  ::::: :::
Db        317 RTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNWTYKKARHBARWDWVWHSAWKK 376

Qy        445 TGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTG 504
              : | | :  :: : :  ::  :      ||:::: :| :   : : ::           : :
Db        377 WHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNGGRTYY 436

Qy        505 AAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGC 564
              : :  : :|  :::|   ::  ::: | :    ::: : :  ::: :  | :: :
Db        437 GWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWKSMTNGKSHRBA 496

Qy        565 ATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTG 624
              | : |:::: : :  |:     :::  :   ::  :  ::::    | : | : :     :
Db        497 AAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWWNNYAAWYTKSSWNYTSRYY 556

Qy        625 TAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGG 684
              ::   ::: : : ::: :   : :  ::| :   :::::: :  : | | ::: :
Db        557 RWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSHTWBHBRAGAAHYWMBMMY 616

Qy        685 GCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACATGTGTGTA 744
              :     : |::: :::||                |   ::::: |  ::     ::
Db        617 BAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNNNNNNATCARDDYYAASRWYAMANAKW 676

Qy        745 TATTGAAAAATCTTGTCTCTTTTTTATT 771
              : :::||   ::|     :  :  ||:
Db        677 YYYKBAANNAYYTHANNWWGCWNNATD 703


RESULT 12
US-09-949-016-142482/c
; Sequence 142482, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 142482
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-142482

  Query Match             5.0%;  Score 39;  DB 4;  Length 601;
  Best Local Similarity   46.5%;  Pred. No. 1;
  Matches  126;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        325 TTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTA 384
              |||| |||    |      ||  | | |||||||  |  |||  ||||     |   ||  |
Db        406 TTTACAAACTTTTAGAATTATCTTATATAAATACTATAACTTTAATACAAAATTTTCAAA 347

Qy        385 ATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATG 444
              | |  |||      |   | |  |||  |||| | |     | | | |  |  || |
Db        346 TTTCACTTACAAAATCTCTGTGGTTACTTCTGAGAAAGGTGTGTGTTTTTTTTTTTTTTT 287
```

```
Qy        445 TGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTG 504
              |         |||    |||  | |  ||||  ||| |  |  ||  | || |||||
Db        286 TTTGGACTAAGTACCAAAATGACCACATTTTTGTGAAAGTGGGCCAAAGTAAAGTTTTAA 227

Qy        505 AAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGC 564
                | ||  ||    |  |   | || ||||| |||  |  |   |||    | | |
Db        226 TTATTATAGTACACCAATTGCATGTTTTGAAACAATCATGTATGAATTAGTAAACTTTTT 167

Qy        565 ATGATTTATGATGAATCATTTTGTGGTCTGT 595
                | | | || || |  ||| |  ||| |
Db        166 TCTTTGTTTTATAAAACCGTTTTTTTTCTTT 136


RESULT 13
US-09-949-016-12372
; Sequence 12372, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12372
;   LENGTH: 181429
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12372

  Query Match             5.0%;  Score 39;  DB 4;  Length 181429;
  Best Local Similarity   46.5%;  Pred. No. 6.6;
  Matches  126;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        325 TTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTA 384
              |||| |||    |      ||  | | |||||||   |  ||| ||||      |   ||   |
Db      11497 TTTACAAACTTTTAGAATTATCTTATATAAATACTATAACTTTAATACAAAATTTTCAAA 11556

Qy        385 ATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATG 444
                | |  |||      |   | |  |||  |||| | |       | | | |  |  || |
Db      11557 TTTCACTTACAAAATCTCTGTGGTTACTTCTGAGAAAGGTGTGTGTTTTTTTTTTTTTTT 11616

Qy        445 TGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTG 504
              |         |||    |||  | |  ||||  ||| |  |  ||  | || |||||
Db      11617 TTTGGACTAAGTACCAAAATGACCACATTTTTGTGAAAGTGGGCCAAAGTAAAGTTTTAA 11676

Qy        505 AAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGC 564
                | ||  ||    |  |   | || ||||| |||  |  |   |||    | | |
Db      11677 TTATTATAGTACACCAATTGCATGTTTTGAAACAATCATGTATGAATTAGTAAACTTTTT 11736

Qy        565 ATGATTTATGATGAATCATTTTGTGGTCTGT 595
                | | | || || |  ||| |  ||| |
Db      11737 TCTTTGTTTTATAAAACCGTTTTTTTTCTTT 11767


RESULT 14
US-09-949-016-15772
; Sequence 15772, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15772
;   LENGTH: 181430
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15772

  Query Match             5.0%;  Score 39;  DB 4;  Length 181430;
  Best Local Similarity   46.5%;  Pred. No. 6.6;
  Matches  126;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        325 TTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTA 384
              |||| |||    |      ||  | | |||||||   |  ||| ||||      |   ||   |
Db      11497 TTTACAAACTTTTAGAATTATCTTATATAAATACTATAACTTTAATACAAAATTTTCAAA 11556

Qy        385 ATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATG 444
                | |  |||      |   | |  |||  |||| | |       | | | |  |  || |
Db      11557 TTTCACTTACAAAATCTCTGTGGTTACTTCTGAGAAAGGTGTGTGTTTTTTTTTTTTTTT 11616

Qy        445 TGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTG 504
              |         |||    |||  | |  ||||  ||| |  |  ||  | || |||||
Db      11617 TTTGGACTAAGTACCAAAATGACCACATTTTTGTGAAAGTGGGCCAAAGTAAAGTTTTAA 11676

Qy        505 AAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGC 564
                | ||  ||    |  |   | || ||||| |||  |  |   |||    | | |
Db      11677 TTATTATAGTACACCAATTGCATGTTTTGAAACAATCATGTATGAATTAGTAAACTTTTT 11736

Qy        565 ATGATTTATGATGAATCATTTTGTGGTCTGT 595
                | | | || || |  ||| |  ||| |
Db      11737 TCTTTGTTTTATAAAACCGTTTTTTTTCTTT 11767


RESULT 15
US-09-679-409-9/c
; Sequence 9, Application US/09679409
; Patent No. 6555316
; GENERAL INFORMATION:
;  APPLICANT:  Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION:  SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE:  53.US15.CIP
;  CURRENT APPLICATION NUMBER: US/09/679,409
;  CURRENT FILING DATE:  2000-10-03
;  PRIOR APPLICATION NUMBER:  09/539,333
;  PRIOR FILING DATE:  2000-03-03
;  PRIOR APPLICATION NUMBER:  09/416,384
;  PRIOR FILING DATE:  1999-10-12
;  PRIOR APPLICATION NUMBER:  60/168,088
;  PRIOR FILING DATE:  1999-11-30
;  NUMBER OF SEQ ID NOS:  134
;  SOFTWARE:  Patent.pm
; SEQ ID NO 9
;   LENGTH: 1783
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: 1..95
;   NAME/KEY: CDS
;   LOCATION: 96..167
;   NAME/KEY: 3'UTR
;   LOCATION: 168..1783
```

; NAME/KEY: polyA_signal
; LOCATION: 1757..1762
; NAME/KEY: allele
; LOCATION: 132
; OTHER INFORMATION: 8-130-143 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 132
; OTHER INFORMATION: 8-126-286 : polymorphic base A or G
US-09-679-409-9

Query Match 4.9%; Score 38.4; DB 4; Length 1783;
Best Local Similarity 45.9%; Pred. No. 2.1;
Matches 168; Conservative 0; Mismatches 196; Indels 2; Gaps 1;

```
Qy        314 TGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTT 373
              | ||   | ||   || |  | |  | ||| | |||          |  || ||| | |||
Db       1535 TAAACAGAAATCCAAATATAGGTGATATTTTCAATGCTCCTCAACTACAAAATTTAATTT 1476

Qy        374 CAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTAT 433
               || | |   | || | || |   |  | ||   | ||  ||   |  |   | | |
Db       1475 TAGAGATGGCTAACCATCTATGAAGTGTTATTGGATGAGCATGGTAACAATATTATGAAA 1416

Qy        434 CGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAG 493
              ||  | | || |||||     |||||  ||       || ||  | |   | |||  |
Db       1415 TAGAGTTTTTCTCACTTAAAATCCCAAAAGGAGCCCTGAACAAGTCATTTTAGAAAAGAA 1356

Qy        494 GAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTG 553
              |||  |      | | || ||   | | |  |     |  || || |  | |||| ||
Db       1355 AAAAAAATACTGAATAATATTTTCACAAGCGTTAAATTAGATATTAAGTATAAATGAAAC 1296

Qy        554 TCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAAT 613
              | ||  | || |  || |   |  ||   |||||  | |     ||||   || |    |
Db       1295 TTCAAACATGTAACATATTAAAGTAAC--TTTTGTTATTTAAGCATTACCAATAAAATTT 1238

Qy        614 AGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTA 673
              |||  ||| |   | ||       ||  ||| || |  | |    |   || || |
Db       1237 CAAAATTATTCTTACTTTAGATTTTAGAAGAATTAAGTAAAAGAGATCAGCCAGGAGTTC 1178

Qy        674 AATTTA 679
              | ||
Db       1177 TGTGTA 1172
```

Search completed: November 8, 2005, 19:52:49
Job time : 199.983 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 18:06:53 ; Search time 1005.4 Seconds
(without alignments)
6391.142 Million cell updates/sec

Title: US-09-463-542-34_COPY_368_1144
Perfect score: 777
Sequence: 1 tcatgtaggtaagactgtgt.....gtctctttttattgttaag 777

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49.2	6.3	18683	16	US-10-311-455-285	Sequence 285, App
2	49.2	6.3	18683	17	US-10-240-452-33	Sequence 33, Appl
3	49	6.3	11422	16	US-10-311-455-192	Sequence 192, App
4	49	6.3	11422	18	US-10-257-166-18	Sequence 18, Appl
5	48.4	6.2	13123	17	US-10-240-452-64	Sequence 64, Appl
c 6	48	6.2	595	13	US-09-925-065A-814225	Sequence 814225,
c 7	48	6.2	1693	14	US-10-027-632-256987	Sequence 256987,
c 8	48	6.2	1693	14	US-10-027-632-256988	Sequence 256988,
c 9	48	6.2	1693	18	US-10-027-632-256987	Sequence 256987,
c 10	48	6.2	1693	18	US-10-027-632-256988	Sequence 256988,
11	48	6.2	3673778	17	US-10-312-841-2	Sequence 2, Appli
12	47	6.0	6880	18	US-10-221-613-183	Sequence 183, App
13	46.8	6.0	6179	19	US-10-221-714A-65	Sequence 65, Appl
14	46.6	6.0	33053	20	US-10-433-793-36	Sequence 36, Appl
15	46.6	6.0	3673778	17	US-10-312-841-1	Sequence 1, Appli
c 16	46.4	6.0	50000	22	US-10-706-635-26	Sequence 26, Appl
17	45.2	5.8	2000	9	US-09-938-842A-5299	Sequence 5299, Ap
18	45.2	5.8	2000	11	US-09-938-842A-5299	Sequence 5299, Ap
19	45.2	5.8	5718	16	US-10-311-455-1346	Sequence 1346, Ap
20	45.2	5.8	5718	19	US-10-221-714A-186	Sequence 186, App
21	44.4	5.7	5886	16	US-10-311-455-2186	Sequence 2186, Ap
22	44.4	5.7	5992	18	US-10-221-613-167	Sequence 167, App
23	44.4	5.7	16811	16	US-10-311-455-1919	Sequence 1919, Ap
24	44.4	5.7	18624	16	US-10-311-455-1675	Sequence 1675, Ap
25	44.2	5.7	655	13	US-09-925-065A-776150	Sequence 776150,
c 26	44.2	5.7	657	13	US-09-925-065A-790118	Sequence 790118,
27	44.2	5.7	5937	16	US-10-240-485-95	Sequence 95, Appl
28	44.2	5.7	6823	16	US-10-311-455-1118	Sequence 1118, Ap
29	44	5.7	13584	16	US-10-311-455-588	Sequence 588, App
30	43.8	5.6	17918	18	US-10-221-613-381	Sequence 381, App
c 31	43.4	5.6	2000	9	US-09-938-842A-3892	Sequence 3892, Ap
c 32	43.4	5.6	2000	11	US-09-938-842A-3892	Sequence 3892, Ap
33	43.4	5.6	6849	19	US-10-240-589C-62	Sequence 62, Appl
34	43.4	5.6	10543	18	US-10-221-613-120	Sequence 120, App
35	43.2	5.6	6056	16	US-10-311-455-999	Sequence 999, App
36	43.2	5.6	7516	16	US-10-311-455-33	Sequence 33, Appl
37	43	5.5	6565	19	US-10-221-714A-188	Sequence 188, App
38	43	5.5	13321	19	US-10-221-714A-144	Sequence 144, App
39	42.8	5.5	6022	19	US-10-221-714A-383	Sequence 383, App
40	42.8	5.5	7072	18	US-10-221-613-347	Sequence 347, App
41	42.8	5.5	12393	16	US-10-311-455-1235	Sequence 1235, Ap
42	42.6	5.5	7676	16	US-10-240-485-152	Sequence 152, App
43	42.6	5.5	21537	16	US-10-311-455-1971	Sequence 1971, Ap
c 44	42.4	5.5	1032	18	US-10-282-122A-16418	Sequence 16418, A
45	42.4	5.5	7384	16	US-10-311-455-719	Sequence 719, App

ALIGNMENTS

RESULT 1
US-10-311-455-285
; Sequence 285, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 285
; LENGTH: 18683
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-285

```
  Query Match             6.3%;  Score 49.2;  DB 16;  Length 18683;
  Best Local Similarity   46.5%;  Pred. No. 0.45;
  Matches  228;  Conservative    0;  Mismatches  258;  Indels    4;  Gaps    2;

Qy        290 TTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATG 349
              |||| ||  |   ||  | |    |   ||| || ||    ||| ||| ||   ||
Db       6355 TTAAAAATAAGTCGTGGTTTGATAATTATAAATAATTAAATGTAGATAGTTATTATTATT 6414

Qy        350 TATAAATTTTCA--AACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTC 407
              |  || |||| |  ||   ||  ||  |  ||  ||||   || ||   | | ||||
Db       6415 TTAAATTTTGAGGAATTGTATTTTGTAATGTGTTTAAGTATTATATTATTTTGTTTGGA 6474

Qy        408 TTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGC 467
              | ||  |       |  | || |  |  || |  |||  |    ||    |
Db       6475 ATTAAAATTGTGGTTTTGATAAGTTTTTTTTTTTAATTGATTTTTGTTATTTTTATTATT 6534

Qy        468 TTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCT 527
              |  |   | ||| | |  ||      ||  |    || | |||| |   ||  || |
Db       6535 ATTGTAGTTTTTATTTAGTGATTATTATAATAATAAAAAATAATTTATAGTAATTATTTT 6594

Qy        528 TTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTG 587
              ||||  ||||| || ||| | ||| ||  |    || ||  ||| ||||  |  ||||
Db       6595 TTTAAAAAATATTTATTATATGATAGTTAATTTATGTATTTTTTTTGATTTAATATTTAA 6654

Qy        588 TGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGC--AAACA 645
              |  |   | |  || || |                |||||||    |||| |    | | |
Db       6655 TATTAATTAATATAGTTGTGTTATTATTTTTATATTGTAGTTGAAGGGGATATTTAGAGA 6714

Qy        646 TTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCT 705
              |    | ||||   |  ||    || ||| ||  |    ||   ||| ||  |
Db       6715 ATGTTTTTAAAATATGTTTTAGGTTAAAAGTTTGTTGAGAGGAGTGGTTTTAGGTTGGTA 6774

Qy        706 AGGCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTT 765
              |  | |  |  | |||    | | | |  | |||  |||||   ||   | | | ||
Db       6775 TGATTTTTAAATTTGTTATTAATTTAATTTTTTGTAAATATTTTTAAATATTTGTTTTG 6834

Qy        766 TTTATTGTTA 775
              |||  |  |||
Db       6835 TTTTTGATTA 6844


RESULT 2
US-10-240-452-33
; Sequence 33, Application US/10240452
; Publication No. US20030162194A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
;  FILE REFERENCE: 5013.1006
;  CURRENT APPLICATION NUMBER: US/10/240,452
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03969
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 33
;   LENGTH: 18683
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-452-33

  Query Match             6.3%;  Score 49.2;  DB 17;  Length 18683;
  Best Local Similarity   46.5%;  Pred. No. 0.45;
  Matches  228;  Conservative    0;  Mismatches  258;  Indels    4;  Gaps    2;

Qy        290 TTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATG 349
              |||| ||  |   ||  | |    |   ||| || ||    ||| ||| ||   ||
Db       6355 TTAAAAATAAGTCGTGGTTTGATAATTATAAATAATTAAATGTAGATAGTTATTATTATT 6414

Qy        350 TATAAATTTTCA--AACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTC 407
              |  || |||| |  ||   ||  ||  |  ||  ||||   || ||   | | ||||
Db       6415 TTAAATTTTGAGGAATTGTATTTTGTAATGTGTTTAAGTATTATATTATTTTGTTTGGA 6474

Qy        408 TTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGC 467
              | ||  |       |  | || |  |  || |  |||  |    ||    |
Db       6475 ATTAAAATTGTGGTTTTGATAAGTTTTTTTTTTTAATTGATTTTTGTTATTTTTATTATT 6534

Qy        468 TTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCT 527
              |  |   | ||| | |  ||      ||  |    || | |||| |   ||  || |
Db       6535 ATTGTAGTTTTTATTTAGTGATTATTATAATAATAAAAAATAATTTATAGTAATTATTTT 6594

Qy        528 TTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTG 587
              ||||  ||||| || ||| | ||| ||  |    || ||  ||| ||||  |  ||||
Db       6595 TTTAAAAAATATTTATTATATGATAGTTAATTTATGTATTTTTTTTGATTTAATATTTAA 6654

Qy        588 TGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGC--AAACA 645
              |  |   | |  || || |                |||||||    |||| |    | | |
Db       6655 TATTAATTAATATAGTTGTGTTATTATTTTTATATTGTAGTTGAAGGGGATATTTAGAGA 6714

Qy        646 TTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCT 705
              |    | ||||   |  ||    || ||| ||  |    ||   ||| ||  |
Db       6715 ATGTTTTTAAAATATGTTTTAGGTTAAAAGTTTGTTGAGAGGAGTGGTTTTAGGTTGGTA 6774

Qy        706 AGGCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTT 765
              |  | |  |  | |||    | | | |  | |||  |||||   ||   | | | ||
Db       6775 TGATTTTTAAATTTGTTATTAATTTAATTTTTTGTAAATATTTTTAAATATTTGTTTTG 6834

Qy        766 TTTATTGTTA 775
              |||  |  |||
Db       6835 TTTTTGATTA 6844


RESULT 3
US-10-311-455-192
; Sequence 192, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 192
;   LENGTH: 11422
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-192

  Query Match             6.3%;  Score 49;  DB 16;  Length 11422;
  Best Local Similarity   46.4%;  Pred. No. 0.41;
  Matches  235;  Conservative    0;  Mismatches  265;  Indels    7;  Gaps    2;

Qy        272 TGTACTGTCACCTTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAA 331
```

```
Db     6619 TTTTTTTAATTTAAATTTAAAAAAATTTTTTTTTTTTAAGAATATATATATTTTTT 6678

Qy      332 AGGATCGTTTTTGCCATGTATAAATTTCAAACATTAACTTTCAGGGTTATTAATCCTTT 391
Db     6679 ATGTTTGTTTTAAGATTTAAAGACGTATATATTATTTTAAAGTAGTGTTATTA-----TT 6733

Qy      392 TAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATA 451
Db     6734 GAAATTTATTATAGTTTTAATTTGTGTTAATTATATAGTTAGAGAAGATAGTAATATAGA 6793

Qy      452 AAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGAT 511
Db     6794 AAATAAGAAATATAATGTTTTAGATATAAGTTATGTAAATTAAAAATTTAGTTTTATTAA 6853

Qy      512 TTTAATAGATGTTTCTTTTATGAAATAATTTTTAA--AAGATTGTCCAGCCCTGCATGAT 569
Db     6854 TTGAATGGATGTTTTTGATATGTATTTTTTTTTAATTTTTATATTTTAATATTTATTTAA 6913

Qy      570 TTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGC 629
Db     6914 GTTATATAAATTTTAGTATTGGTAATTTTTTATTTAGAAGATTTTGTAGGTAGTTATATA 6973

Qy      630 TCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATG 689
Db     6974 TGATGAGAAGTTATAATTTTTTGGTTAAGATAGGAAAAATAAAATGAATAATTTTTATTA 7033

Qy      690 TACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTG 749
Db     7034 AATTTTATAAAAGAGATATATTTGTTAATTTGTAGTGTATATGTTTGGAGAATATATTT 7093

Qy      750 AAAATCTTGTCTCTTTTTTATTGTTAA 776
Db     7094 ATGTTTTTGGCGTTTATTTTTTTTGAA 7120
```

RESULT 4
US-10-257-166-18
; Sequence 18, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 18
; LENGTH: 11422
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-18

```
  Query Match             6.3%;  Score 49;  DB 18;  Length 11422;
  Best Local Similarity   46.4%;  Pred. No. 0.41;
  Matches  235;  Conservative    0;  Mismatches  265;  Indels    7;  Gaps    2;

Qy      272 TGTACTGTCACCTTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAA 331
Db     6619 TTTTTTTAATTTAAATTTAAAAAAATTTTTTTTTTTTAAGAATATATATATTTTTT 6678

Qy      332 AGGATCGTTTTTGCCATGTATAAATTTCAAACATTAACTTTCAGGGTTATTAATCCTTT 391
```

RESULT 5
US-10-240-452-64
; Sequence 64, Application US/10240452
; Publication No. US20030162194A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
; FILE REFERENCE: 5013.1006
; CURRENT APPLICATION NUMBER: US/10/240,452
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03969
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 64
; LENGTH: 13123
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (8274..)
US-10-240-452-64

```
  Query Match             6.2%;  Score 48.4;  DB 17;  Length 13123;
  Best Local Similarity   49.1%;  Pred. No. 0.61;
  Matches  157;  Conservative    0;  Mismatches  161;  Indels    2;  Gaps    1;

Qy      284 TTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTT 343
Db     2473 TTATGTATAAAATTTTATTTAATTTAGTTTAAATTTTATTTTAAAAGTATTTTTGTT 2532
```

```
Qy        344 GCCATGTATAAATTTTCAAACATTAACTT--TCAGGGTTATTAATCCTTTTAAGGTCTAG 401
              | ||| | || |   |||   ||  |  | |      |  |||||  |  |
Db       2533 TTTTGTTTTAATTATTTATGTATTTTTTTAGTTTTGATAGCGTTTAGTTTTATTGATTTT 2592

Qy        402 TTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAA 461
              ||||| ||   |      | |||||  |||||     ||  | | |  | |   ||  |
Db       2593 TTTTTATTGTTTTGAAAGATTAATATGGGTATTAATTTTTTTATAAGTTTATTTATTTAT 2652

Qy        462 AGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGAT 521
              |  ||| || |||     | |   |||| |  ||       |  ||| | |  ||
Db       2653 TAAAAATATATTTATATTATAAGTTGTTTTAGGTATTTATTAAGTATTTTTTAAGTTTAT 2712

Qy        522 GTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATC 581
              ||| ||||||  | |   ||| ||   |||||          ||  ||| | ||   |
Db       2713 TTTTTTTTTATTTATTGTATTTGAATTAATTGTTAGAGTTAAGATATTTTTTTATTTTTT 2772

Qy        582 ATTTTGTGGTCTGTTAGTTA 601
              |||| ||||  |||   ||
Db       2773 TTTTTTTGGTTAGTTTTGTA 2792
```

RESULT 6
US-09-925-065A-814225/c
; Sequence 814225, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 814225
; LENGTH: 595
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-814225

Query Match 6.2%; Score 48; DB 13; Length 595;
Best Local Similarity 46.2%; Pred. No. 0.2;
Matches 159; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

```
Qy        411 AGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTC 470
              ||   ||| |  | | |||| |  ||  | ||    |||   || |||||      ||
Db        452 AGCAGGTGGAAAATTTGAGGAAGTGTGTTGCAATCTACAGTGAACATGGACTATTTCTAG 393

Qy        471 ATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTT 530
              |   | |      || |  |||         | |  || |  | |  | |   |||| ||
Db        392 AGCGAAGAAGTATATTGGGATAAATGTTAGTGAGTGGTAAAGTGAGCACAGAGTTCTATT 333

Qy        531 ATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGG 590
              || || ||||||||  |||| |   |||  |   |||    ||     |    | || ||
Db        332 ATTAAGTAATTTTTGCAAGACTTATCAGGGCAAAATGTCAGATTGCAATGGGTATTCTGA 273

Qy        591 TCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAG 650
              |||  ||| |      | | | | |  ||| ||      |   |    ||     |
Db        272 ACTGAAAGTGAAAAAGTTAAACTGGGAGTCATAGTCTAAATATCTATTTTAATATATTTA 213

Qy        651 AATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCT 710
                |||||    ||| || | | | |||    |       ||||| ||       | |  |
```

```
Db        212 TCTGAAAATGTATAAAGGATGGATATTAGACTCCCATGTTACATGTTACTTATTTCGTGT 153

Qy        711 GTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAAT 754
              || |      | ||  | | ||||| || ||||||| | |
Db        152 TAATATAAAGAAGTAAATCTTAAGCATGTATGAATATTGAGAGT 109
```

RESULT 7
US-10-027-632-256987/c
; Sequence 256987, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 256987
; LENGTH: 1693
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-256987

Query Match 6.2%; Score 48; DB 14; Length 1693;
Best Local Similarity 46.2%; Pred. No. 0.31;
Matches 159; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

```
Qy        411 AGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTC 470
              ||   ||| |  | | |||| |  ||  | ||    |||   || |||||      ||
Db        475 AGCAGGTGGAAAATTTGAGGAAGTGTGTTGCAATCTACAGTGAACATGGACTATTTCTAG 416

Qy        471 ATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTT 530
              |   | |      || |  |||         | |  || |  | |  | |   |||| ||
Db        415 AGCGAAGAAGTATATTGGGATAAATGTTAGTGAGTGGTAAAGTGAGCACAGAGTTCTATT 356

Qy        531 ATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGG 590
              || || ||||||||  |||| |   |||  |   |||    ||     |    | || ||
Db        355 ATTAAGTAATTTTTGCAAGACTTATCAGGGCAAAATGTCAGATTGCAATGGGTATTCTGA 296

Qy        591 TCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAG 650
              |||  ||| |      | | | | |  ||| ||      |   |    ||     |
Db        295 ACTGAAAGTGAAAAAGTTAAACTGGGAGTCATAGTCTAAATATCTATTTTAATATATTTA 236

Qy        651 AATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCT 710
                |||||    ||| || | | | |||    |       ||||| ||       | |  |
Db        235 TCTGAAAATGTATAAAGGATGGATATTAGACTCCCATGTTACATGTTACTTATTTCGTGT 176

Qy        711 GTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAAT 754
              || |      | ||  | | ||||| || ||||||| | |
Db        175 TAATATAAAGAAGTAAATCTTAAGCATGTATGAATATTGAGAGT 132
```

RESULT 8
US-10-027-632-256988/c
; Sequence 256988, Application US/10027632
; Publication No. US20020198371A1

```
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 256988
;   LENGTH: 1693
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-256988

  Query Match             6.2%;  Score 48;  DB 14;  Length 1693;
  Best Local Similarity   46.2%;  Pred. No. 0.31;
  Matches  159;  Conservative    0;  Mismatches  185;  Indels    0;  Gaps    0;

Qy        411 AGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTC 470
              ||   ||| |  | | |||| |  ||  | ||     |||    || |||||        ||
Db        475 AGCAGGTGGAAAATTTGAGGAAGTGTGTTGCAATCTACAGTGAACATGGACTATTTCTAG 416

Qy        471 ATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTT 530
              |    | |       || |  |||          | |  || |  | |  | |   |||| ||
Db        415 AGCGAAGAAGTATATTGGGATAAATGTTAGTGAGTGGTAAAGTGAGCACAGAGTTCTATT 356

Qy        531 ATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGG 590
              || || ||||||||  |||| |   |||  |    ||     |    | || ||
Db        355 ATTAAGTAATTTTTGCAAGACTTATCAGGGCAAAATGTCAGATTGCAATGGGTATTCTGA 296

Qy        591 TCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAG 650
              |||  ||| |       | | | |  ||| ||     |    |    ||    |
Db        295 ACTGAAAGTGAAAAAGTTAAACTGGGAGTCATAGTCTAAATATCTATTTTAATATATTTA 236

Qy        651 AATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCT 710
              |||||    ||| || | | | |||    |        |||| ||        | | |
Db        235 TCTGAAAATGTATAAAGGATGGATATTAGACTCCCATGTTACATGTTACTTATTTYGTGT 176

Qy        711 GTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAAT 754
                || |      | ||  | | |||| || |||||| | |
Db        175 TAATATAAAGAAGTAAATCTTAAGCATGTATGAATATTGAGAGT 132


RESULT 9
US-10-027-632-256987/c
; Sequence 256987, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 256987
;   LENGTH: 1693
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-256987

  Query Match             6.2%;  Score 48;  DB 18;  Length 1693;
  Best Local Similarity   46.2%;  Pred. No. 0.31;
  Matches  159;  Conservative    0;  Mismatches  185;  Indels    0;  Gaps    0;

Qy        411 AGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTC 470
              ||   ||| |  | | |||| |  ||  | ||     |||    || |||||        ||
Db        475 AGCAGGTGGAAAATTTGAGGAAGTGTGTTGCAATCTACAGTGAACATGGACTATTTCTAG 416

Qy        471 ATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTT 530
              |    | |       || |  |||          | |  || |  | |  | |   |||| ||
Db        415 AGCGAAGAAGTATATTGGGATAAATGTTAGTGAGTGGTAAAGTGAGCACAGAGTTCTATT 356

Qy        531 ATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGG 590
              || || ||||||||  |||| |   |||  |    |||    ||     |    | || ||
Db        355 ATTAAGTAATTTTTGCAAGACTTATCAGGGCAAAATGTCAGATTGCAATGGGTATTCTGA 296

Qy        591 TCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAG 650
              |||  ||| |       | | | |  ||| ||     |    |    ||    |
Db        295 ACTGAAAGTGAAAAAGTTAAACTGGGAGTCATAGTCTAAATATCTATTTTAATATATTTA 236

Qy        651 AATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCT 710
              |||||    ||| || | | | |||    |        |||| ||        | | |
Db        235 TCTGAAAATGTATAAAGGATGGATATTAGACTCCCATGTTACATGTTACTTATTTCGTGT 176

Qy        711 GTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAAT 754
                || |      | ||  | | |||| || |||||| | |
Db        175 TAATATAAAGAAGTAAATCTTAAGCATGTATGAATATTGAGAGT 132


RESULT 10
US-10-027-632-256988/c
; Sequence 256988, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 256988
;   LENGTH: 1693
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-256988

  Query Match             6.2%;  Score 48;  DB 18;  Length 1693;
  Best Local Similarity   46.2%;  Pred. No. 0.31;
  Matches  159;  Conservative    0;  Mismatches  185;  Indels    0;  Gaps    0;

Qy        411 AGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTC 470
              ||   ||| |  | | |||| |  ||  | ||    |||   || |||||      ||
Db        475 AGCAGGTGGAAAATTTGAGGAAGTGTGTTGCAATCTACAGTGAACATGGACTATTTCTAG 416

Qy        471 ATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTT 530
              |   | |       || |  |||          | |  || |  | |  | |   |||| ||
Db        415 AGCGAAGAAGTATATTGGGATAAATGTTAGTGAGTGGTAAAGTGAGCACAGAGTTCTATT 356

Qy        531 ATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGG 590
              || || |||||||  |||| |   |||  |   |||    ||    |    | || ||
Db        355 ATTAAGTAATTTTTGCAAGACTTATCAGGGCAAAATGTCAGATTGCAATGGGTATTCTGA 296

Qy        591 TCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAG 650
              |||  ||| |       | | | | |  ||| ||      |   |    ||    |
Db        295 ACTGAAAGTGAAAAAGTTAAACTGGGAGTCATAGTCTAAATATCTATTTAATATATTTA 236

Qy        651 AATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCT 710
              |||||    ||| || | | |||    |       |||| ||       | |  |
Db        235 TCTGAAAATGTATAAAGGATGGATATTAGACTCCCATGTTACATGTTACTTATTTYGTGT 176

Qy        711 GTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAAT 754
                || |       | ||  | | |||| || |||||| | |
Db        175 TAATATAAAGAAGTAAATCTTAAGCATGTATGAATATTGAGAGT 132


RESULT 11
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (379615)
US-10-312-841-2

  Query Match             6.2%;  Score 48;  DB 17;  Length 3673778;
  Best Local Similarity   49.4%;  Pred. No. 9;
  Matches  153;  Conservative    0;  Mismatches  155;  Indels    2;  Gaps    1;

Qy        323 ATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTAT 382
              ||| |||| |   |  ||||||  || | | | ||||      ||   |||  |  ||||
Db     717761 ATTGTAAATATTTTTTTTTTTGGGATTTGTTATTTTTTTTTTGTTTTGTTTTGGTTTTAT 717820

Qy        383 TAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCA 442
              |   || |||  | ||||| || ||  ||  |    || |   || |||  |  ||
Db     717821 TTTTGTTTGTAATTTTTAGTTATTTTTTTGTTAGGTTTGTGAGTAAGTTATTTTTGTTAG 717880

Qy        443 TGTGACATAAAAGATGGAAAGGGGCTTCAT--TCATGTTAGTGATGGAAATAGGAAAGTA 500
              |          |  ||| | |   || ||  | ||      |   |    ||||    |
Db     717881 TTGTGTTATTTATTAGGATATGATTTTTATGGTGATTGAGAAGTGGTTGTTAGGTTGTTG 717940

Qy        501 GGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTTAAAAGATTGTCCAGCC 560
              ||   |  ||||  |  |  |||  ||| | | | | |||||||| | ||||  |
Db     717941 TTTGTTTTTGTTTTTTTTTAAATTTTTGTTTGTAAGAAATTTTTTTAAATGTTTGTTTATTA 718000

Qy        561 CTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGC 620
              |     | | |   |      ||||||  || ||||  ||   ||| ||  |||   |
Db     718001 TTATGATAATGAGAGTTTTGTTTTTTGTTGTTTGTTGTTTTTGTTTGGATTATAAGTATT 718060

Qy        621 ATTGTAGGCT 630
              || || | |
Db     718061 TTTATAAGTT 718070


RESULT 12
US-10-221-613-183
; Sequence 183, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, ALexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
;  FILE REFERENCE: 5013.1004
;  CURRENT APPLICATION NUMBER: US/10/221,613
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: PCT/EP01/02945
;    DE 10013847.00
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-03-15
;    2000-03-15
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 183
;   LENGTH: 6880
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-183

  Query Match             6.0%;  Score 47;  DB 18;  Length 6880;
  Best Local Similarity   51.8%;  Pred. No. 1;
  Matches  131;  Conservative    0;  Mismatches  120;  Indels    2;  Gaps    1;

Qy        301 GTGTCATCTAGAATGAAGTTATATTTTAAAA--AGGATCGTTTTTGCCATGTATAAATTT 358
              |||| || | |  || || ||  | ||||||   ||    |||||   | ||||||||
Db        722 GTGTTATTTTGTGTGTAGATAGTTGTTAAAATTTGGTGTTTTTTTTAGAGGTATAAATGA 781

Qy        359 TCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTG 418
              |  |   ||   |||      || ||  |  ||||    | | ||||||    ||    |
Db        782 TGTAGGTTTTTATTTTGTTATTTTTTTTTGTTTTGTTTTTTTTGTTTTTTGAGAGGTATTA 841

Qy        419 CAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGT 478
              | |||| ||  | |  |   | || |   || |     |   || |   ||  ||  |||
Db        842 TATTAATGGAAATTTAATATATAATTTATAATTATGTTTTTGAAAGTTTTTTTTTTTTGT 901

Qy        479 TAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATA 538
              |  | ||    | |    ||||| |  || | | | | | | ||  || |||| |  |
Db        902 TTTTAATATTTAAATTTTAGTAGTTTTAGGGGTATAAGTGGTTTTTGATTATATGGATGA 961

Qy        539 ATTTTTAAAAGAT 551
              ||| |  | | ||
Db        962 ATTGTATAGAAAT 974


RESULT 13
```

```
US-10-221-714A-65
; Sequence 65, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 65
;   LENGTH: 6179
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-65

  Query Match             6.0%;  Score 46.8;  DB 19;  Length 6179;
  Best Local Similarity   53.0%;  Pred. No. 1.1;
  Matches  122;  Conservative    0;  Mismatches  107;  Indels    1;  Gaps    1;

Qy        319 TTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGG 378
              |||| |||| ||| |  ||| ||   ||  ||||| |||    |  || |   |  ||
Db       3255 TTATTGTTTAGTAAGTAATGTTATTAGAATAGATAAAGTTTGGTA-TTTGAGAATATGGA 3313

Qy        379 TTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCA 438
              ||| ||     | |  | |||| ||| |||| | |||        ||  |  ||  |
Db       3314 ATATATATTAAGGTGATATATAGTGTTTATTAAATGTGTTGTAATGTAATGTGATAATTT 3373

Qy        439 TTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAG 498
              |    | |  | ||||  || ||  ||  ||  |  || ||  ||   |   ||  || |
Db       3374 TAATAGAGGAAAAAAATGTGAAATTGGATTTGTTATATATTTTTGTATGTTTTAAAAATG 3433

Qy        499 TAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAA 548
              |||   |  |  || | |||  | ||| ||||| | |    ||||||  ||
Db       3434 TAGTATAGATTTTTGTTATATTTATTTATTTTAAGTATATTTTTTTATAA 3483


RESULT 14
US-10-433-793-36
; Sequence 36, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/433,793
;  CURRENT FILING DATE:  2003-06-06
;  NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 36
;   LENGTH: 33053
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-36

  Query Match             6.0%;  Score 46.6;  DB 20;  Length 33053;
  Best Local Similarity   46.7%;  Pred. No. 2.6;
  Matches  148;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy        290 TTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATG 349
              || ||  |  || | | || ||| || ||||  |||     | |     |  | |||   | |
Db      14921 TTTATGATTATTTTTTATTTAGTATAAAGTATTATAAAGAGATTAGTTTTGTTTTATAGG 14980

Qy        350 TATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTT 409
              ||||  |  |    ||    |    ||  |  |||       |  |||||   |  | |||||  |
Db      14981 AATAATTATATTAATGATGTTTTATAATGTTTGATGTAATTTTAGTTTTAATTTTTTTGT 15040

Qy        410 AAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTT 469
              |  |  |    |  | | | ||  ||  |||||  | |  |
Db      15041 TATACGTTTGTTGTGAAGGTTGTGGTGTTTTTTGTGATTTTATTAAGAATGTTTTTTGG 15100

Qy        470 CATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTT 529
              | |||||  |  | | | ||     |  ||  |     | |       | |||  |||
Db      15101 TTGATGTTTTAGTAAATGTAGTGGGTGTATTATTGTTGAATATATTTATATTATTTTTTT 15160

Qy        530 TATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTG 589
              |||  |||| ||||| | | ||      ||    |   ||||| |||  | |  | ||| |
Db      15161 TATAGAATATTTTTTGATATATAAAAAAGTTATTTTTGATTTTTGAGAATTTTTATTGAG 15220

Qy        590 GTCTGTTAGTTACTTTT 606
                  ||  ||  | ||
Db      15221 AAAATTTTTTTTTTGTT 15237


RESULT 15
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             6.0%;  Score 46.6;  DB 17;  Length 3673778;
  Best Local Similarity   46.7%;  Pred. No. 20;
  Matches  148;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy        307 TCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACAT 366
              | || | |  || || ||  ||| | ||   ||| |  |  | |   || |  |  ||
Db    1671903 TTTATATTTATTTTTTAGTTGGAAAGGAATGTATTTAGTAAATTTTCGGTTGTATATTAT 1671962

Qy        367 TAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATA 426
              |    ||  ||  ||||||||  || | ||    |  | |||  | ||| ||      |   |
Db    1671963 TGGTTTGTAGTTTTATTAATTTTTGTTAGTAGTGATATTTAGTTAGTATGGTAGTTGTAA 1672022

Qy        427 GAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATG 486
                  ||  | |||     | |  |  |||          |  ||  ||   ||  |  |
Db    1672023 TTAAGATTATTATTTGGTTAAGTTTGAAGTAATTTTGTTTATTTTTTAGGTTTTATTGGG 1672082

Qy        487 GAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAA 546
              |    |   |||   |   |     ||||||  |  ||  |  |||  |  | ||||||  |
Db    1672083 TTTTTGTAAGGATAGAATATACGGAGATAATAGGTAATTTTAGTATTATTTTATTTTTTA 1672142

Qy        547 AAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTT 606
              || || |  |     |  | |||| | || |    | |  |  | | | || |   |||
```

```
Db    1672143 TAGTTTTTTAATGGTGGTGTTATTTTTTATAATATTTGTAATAATTTTTAAGATTTGTTT 1672202

Qy        607 AGAGAATAGAAAGCATT 623
              | ||    | | | ||
Db    1672203 TGGGACGTAATATCGTT 1672219
```

Search completed: November 8, 2005, 23:53:38
Job time : 1012.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 22:16:08 ; Search time 4014.41 Seconds
(without alignments)
7367.447 Million cell updates/sec

Title: US-09-463-542-34_COPY_368_1144
Perfect score: 777
Sequence: 1 tcatgtaggtaagactgtgt.....gtctctttttattgttaag 777

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	101.6	13.1	376	2	BF736184	BF736184 PM4-KT000
c	2	63.4	8.2	358	9	CC774457	CC774457 CH240_95D
c	3	62.4	8.0	390	1	AU234451	AU234451 AU234451
	4	62.4	8.0	400	1	AU278852	AU278852 AU278852
c	5	58.2	7.5	1101	9	CNS0039G	AL063921 Drosophil
c	6	54.6	7.0	919	9	CNS005RL	AL061409 Drosophil
c	7	53.8	6.9	1101	9	CNS00EVL	AL069706 Drosophil
	8	53.8	6.9	1101	9	CNS00LT2	AL078714 Drosophil
	9	53.8	6.9	1101	9	CNS016LI	AL106896 Drosophil
c	10	52.4	6.7	1200	9	CNS016CO	AL106578 Drosophil
	11	52	6.7	1146	9	CNS021G2	AL176843 Tetraodon
	12	51.8	6.7	1530	9	AG382019	AG382019 Mus muscu
c	13	51.6	6.6	500	1	AU088479	AU088479 AU088479
	14	51.6	6.6	1201	9	CNS016BY	AL106552 Drosophil
c	15	51.4	6.6	945	9	CNS04D0K	AL285149 Tetraodon
c	16	50.2	6.5	1190	9	CNS02ON7	AL206908 Tetraodon
c	17	49.6	6.4	1101	9	CNS00GCX	AL072030 Drosophil
	18	48.8	6.3	928	9	CNS00DKY	AL071865 Drosophil
c	19	48.6	6.3	996	9	CNS00FUH	AL071063 Drosophil
	20	48.4	6.2	1001	9	CNS0140O	AL103554 Drosophil
	21	48.4	6.2	1101	9	CNS0100X	AL098379 Drosophil
	22	48.4	6.2	1201	9	CNS0162X	AL106227 Drosophil
	23	48.2	6.2	1101	9	CNS001FB	AL060732 Drosophil
	24	47.8	6.2	842	8	BZ696583	BZ696583 SP__Ba008
	25	47.6	6.1	979	9	CNS0161W	AL106190 Drosophil
c	26	47.6	6.1	1315	3	CR657214	CR657214 Tetraodon
	27	47.4	6.1	813	9	BX218284	BX218284 Danio rer
c	28	47.4	6.1	1092	9	CNS020K7	AL175696 Tetraodon
	29	47.4	6.1	1169	9	CNS06KHQ	AL402900 T3 end of
	30	47.2	6.1	1101	9	CNS000D1	AL065414 Drosophil
	31	47	6.0	1065	8	BZ696860	BZ696860 SP__Ba008
c	32	47	6.0	1094	9	CNS012FZ	AL101513 Drosophil
	33	47	6.0	1101	9	CNS00Z65	AL097271 Drosophil
	34	46.8	6.0	510	1	AV778356	AV778356 AV778356
c	35	46.8	6.0	609	9	CNS025K2	AL182171 Tetraodon
	36	46.8	6.0	1031	9	CNS00CF2	AL059199 Drosophil
c	37	46.8	6.0	1101	9	CNS003BD	AL064091 Drosophil
c	38	46.8	6.0	1101	9	CNS00ESI	AL069797 Drosophil
	39	46.6	6.0	792	9	CG116113	CG116113 PUJDS79TB
c	40	46.6	6.0	870	9	CG209274	CG209274 OG1EB38TV
c	41	46.6	6.0	878	9	CNS0187R	AL108993 Drosophil
c	42	46.6	6.0	894	8	BZ687506	BZ687506 PUBHR65TD
	43	46.6	6.0	906	9	CG209262	CG209262 OG1EB38TH
c	44	46.4	6.0	1896	9	CG753083	CG753083 P048-1-C0
	45	46	5.9	516	5	BM883909	BM883909 rc24b11.y

ALIGNMENTS

RESULT 1
BF736184/c
LOCUS BF736184 376 bp mRNA linear EST 10-JAN-2001
DEFINITION PM4-KT0005-151100-001-b03 KT0005 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF736184
VERSION BF736184.1 GI:12062858
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 376)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM4&t2=PM4-KT0005-151100-001-b03&t3=2000-11-15&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 21
High quality sequence stop: 376.
FEATURES Location/Qualifiers
source 1. .376
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="KT0005"
/note="Organ: bladder_tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

```
  Query Match             13.1%;  Score 101.6;  DB 2;  Length 376;
  Best Local Similarity   95.8%;  Pred. No. 8.4e-14;
  Matches  115;  Conservative    0;  Mismatches    4;  Indels    1;  Gaps    1;

Qy          1 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TCATGTAGGTAAGACTGTGTAGAATGTCGGGTCTCGATGTTGGCGCTATTCAAGCCCTGA 63

Qy         61 TGATAAGGCTTTTGGCATTAGATGCTGTTTTGTCTTCATGGAAAATACAGCTATTCTAGG 120
              |||||||||||||| |||||||||||||||||||||||| | | |||||||||| |||
Db         62 TGATAAGGCTTTTGTCATTAGATGCTGTTTTGTCTTCATGG-ATACACAGCTATTCGAGG 4
```

RESULT 2
CC774457/c
LOCUS CC774457 358 bp DNA linear GSS 27-JUN-2003
DEFINITION CH240_95D2.TV CHORI-240 Bos taurus genomic clone CH240_95D2, genomic survey sequence.
ACCESSION CC774457
VERSION CC774457.1 GI:32326748
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 358)
AUTHORS Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P., Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L., Womack,J.E., de Jong,P.J. and Lewin,H.A.
TITLE Bovine BAC end sequences from CHORI-240 library
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: CH240_95D2.TJ
Contact: Harris Lewin
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Clones are derived from the bovine BAC library CHORI-240 (http://www.chori.org/bacpac/bovine240.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/ordering_information.htm). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by by University of Illinois at Urbana Champaign, USA with funds provided by grant No. AG202-34480-11828 from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock Genome Sequencing Initiative)
Plate: 95 row: D column: 2
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .358
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_95D2"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI; Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"

ORIGIN

```
  Query Match              8.2%;  Score 63.4;  DB 9;  Length 358;
  Best Local Similarity   71.5%;  Pred. No. 0.00014;
  Matches   98;  Conservative    0;  Mismatches   36;  Indels    3;  Gaps    1;

Qy        283 CTTAGGCTTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTT 342
              ||| | ||| | |||||| | ||||||| | || |||||| | ||   || | ||
Db        140 CTTTAGGTTAGCAGGCATGTGCCCTCTAGAACAAGCTTGTATTTTCATAAATGTCATCTT 81

Qy        343 TGCCATGTATAAATTTTCAAACAT---TAACTTTCAGGGTTATTAATCCTTTTAAGGTCT 399
              ||| ||||||| |||| | || |||| ||||| |||||||||||||| |||
Db         80 TGCTGTGTATAACTTTTATCAATCACCTATCTTTTGAGGTTAGTAATCCTTTTAAGATCT 21

Qy        400 AGTTTTTCTTAAGTCTG 416
              |||||||||||||||||
Db         20 AGTTTTTCTTAAGTCTG 4
```

RESULT 3
AU234451/c
LOCUS AU234451 390 bp mRNA linear EST 21-SEP-2001
DEFINITION AU234451 Bovine placenta cDNA Bos taurus cDNA clone Cln1153 3', mRNA sequence.
ACCESSION AU234451
VERSION AU234451.1 GI:15719669
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 390)
AUTHORS Gohma,H., Lejukole,H.Y., Taniguchi,Y., Yamada,T., Akagi,S., Yasue,H. and Sasaki,Y.
TITLE Analysis of expressed sequence tags from a cDNA library of bovine placenta
JOURNAL Unpublished (2001)
COMMENT Contact: Takahisa Yamada
Graduate School of Agriculture
Kyoto University
Sakyoku, Kitashirakawa, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-6323
Fax: 81-75-753-6340
Email: tyamada@jkans.jkans.kais.kyoto-u.ac.j
This clone was obtained from a 3' end cDNA library.
FEATURES Location/Qualifiers
source 1. .390
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="Cln1153"
/tissue_type="placenta"
/clone_lib="Bovine placenta cDNA"

ORIGIN

```
  Query Match              8.0%;  Score 62.4;  DB 1;  Length 390;
  Best Local Similarity   69.1%;  Pred. No. 0.00025;
  Matches  123;  Conservative    0;  Mismatches   36;  Indels   19;  Gaps    2;

Qy        600 TACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATC 659
              || || ||||||| | ||||  || || ||||||||||||||||||||||  ||||||||
Db        299 TAGTTCTAGAGAACACAAAGTGTTTTAAGCTCAGGGAAAGCAAACATTCAAGATGAAATC 240

Qy        660 CAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATG 719
              |||             || || ||  || | | ||||||| |||   ||  ||| ||  |
Db        239 CAAC-----------ATGTACTTCTGCTTATCCATTTTGACAGATGAGTTTGTCTAAGT- 192

Qy        720 TGTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777
                  ||| |||| |||||||||||||| |||||||| | ||||||   ||||| ||
Db        191 -------ATTTTGAATATGTGTGTATATTAAAAATCTTCTTTCTTTTCAGTTGTTCAG 141
```

```
RESULT 4
AU278852
LOCUS       AU278852                400 bp    mRNA    linear   EST 02-JUL-2002
DEFINITION  AU278852 Cloned bovine placenta cDNA Bos taurus cDNA clone
            placenta1153 3', mRNA sequence.
ACCESSION   AU278852
VERSION     AU278852.1  GI:21682162
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Oishi,M., Yamada,T., Goma,H., Lejukole,H.Y., Taniguchi,Y. and
            Sasaki,Y.
  TITLE     EST analysis of cloned bovine fetus and placenta
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Masahito Oishi
            Graduate School of Agriculture
            Kyoto University
            Sakyoku Kitashirakawa, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-6331
            Fax: 81-75-753-6340
            Email: oishi@jkans.jkans.kais.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="placenta1153"
                     /dev_stage="60 embryonic day"
                     /clone_lib="Cloned bovine placenta cDNA"
ORIGIN

  Query Match             8.0%;  Score 62.4;  DB 1;  Length 400;
  Best Local Similarity   69.1%;  Pred. No. 0.00025;
  Matches  123;  Conservative    0;  Mismatches   36;  Indels   19;  Gaps    2;

Qy        600 TACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATC 659
              || || ||||||| | ||||  || || ||||||||||||||||||||||  ||||||||
Db         92 TAGTTCTAGAGAACACAAAGTGTTTTAAGCTCAGGGAAAGCAAACATTCAAGATGAAATC 151

Qy        660 CAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATG 719
              |||            || || ||  || | | ||||||| |||   ||  ||| || |
Db        152 CAAC----------ATGTACTTCTGCTTATCCATTTTGACAGATGAGTTTGTCTAAGT- 199

Qy        720 TGTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTTAAG 777
                  ||| |||| |||||||||||||| |||||||| | ||||||   ||||| ||
Db        200 -------ATTTTGAATATGTGTGTATATTAAAAATCTTCTTTCTTTTCAGTTGTTCAG 250


RESULT 5
CNS0039G/c
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             7.5%;  Score 58.2;  DB 9;  Length 1101;
  Best Local Similarity   14.3%;  Pred. No. 0.0031;
  Matches   66;  Conservative  223;  Mismatches  172;  Indels    0;  Gaps    0;

Qy        309 TAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTA 368
              |:  || :: ::::||::|::: :  :    :|:|   :::||::|::    : |   ::
Db       1049 TWWTATWWWWWWWWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDR 990

Qy        369 ACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGA 428
              :  |: :| |  :  :   :::  :|:: : | ::::  :||:: |::  : :::::
Db        989 DAATDADGAGRRDGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWK 930

Qy        429 GGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGA 488
              : |: :: :::   |::: :    : ::: :  :::: :: ::: :: :    :||  :
Db        929 WDGAKDRKADDDDGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAW 870

Qy        489 AATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAA 548
              |||:  ::   : ::  ::  |::::|::: :   :: :: :|:||:  :: :: ::|
Db        869 AATDWWWWGWADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGK 810

Qy        549 GATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAG 608
              :   ::  ::   :: :: |:: :: ||   :  ||| :|::: ::|:  | :  : |::
Db        809 RGGARKRRDRKRADDKRDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWD 750

Qy        609 AGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGA 668
              :::::::::|| |  :::  : : :::: : ::| |:: |::: ::: :  |:  : :
Db        749 RDDDDDRDRAGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKK 690

Qy        669 AGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACAT 728
              |:::: : ||::: : :: :| ::|||:   :  : ::  ::| : : :|  : | ::
Db        689 TGRKRRRRDRATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDW 630

Qy        729 TCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTA 769
              : ::|: : ::::::|: : ::::||: :|:  : : :: :|
Db        629 WTWKAWDWAKWDWKTRADRWDRWAADTWTDARKADRDWAKA 589


RESULT 6
CNS005RL/c
LOCUS       CNS005RL                919 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR12F23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL061409
VERSION     AL061409.1  GI:4943512
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
```

```
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 919)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .919
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR12F23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             7.0%;  Score 54.6;  DB 9;  Length 919;
  Best Local Similarity   22.5%;  Pred. No. 0.022;
  Matches   86;  Conservative  128;  Mismatches  168;  Indels    0;  Gaps    0;

Qy        299 ATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTT 358
              |:::::  : :|:| : |  :: ::: ::|: : :: : :: ::::  |: ::  ||::: 
Db        912 AWKKDKKKKDKADAKDAAKAWKKDDDAWKAWKWKKKKKKKKKGKKKKADAKAKRAKAAKKK 853

Qy        359 TCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTG 418
                :  :: | |     | :: :      |::| :  |  : |:|:| |:|:|: ||||
Db        852 AKDGDWKKDKAGTGAKTTGTKKTKGTKGGWTKWTTDTWTWAWTTWTKTATDAWGWATGTG 793

Qy        419 CAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGT 478
              ::  : : :| |   :  :: :| : |  :: : :|:  :::   : :|       :::
Db        792 WTDKGKKGKKKGAAATAKKGWKTWTTWAAWTWDKWTKAWAAWRDAATRTWTDTAAGKKKK 733

Qy        479 TAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATA 538
              : :: |    | :|: | : : ::::: :::::||::  | :| :|| |||| |    :
Db        732 KTKKWAATWGATKAKWATWTKKKKKKDTKWKWKTTWDTNAWWTTWTTTTTTTTTTTTTWT 673

Qy        539 ATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAG 598
              |||||  ||: :|| |          ::   | ::     | || ||||  :|:: |
Db        672 TTTTTTTAAWAWTTTTGTTTTTTATATGTDKAAGTAWWTTTAAAATTNGTGGNGKGKKKG 613

Qy        599 TTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAAT 658
              ::  :::  | :    :   :  ::|: :: : :: ::::  : : ||  :    ||||
Db        612 KKGKKKKGTGKKKKGKKGKKKTKKKGKKKDKKTDDTDDDWWWWTWWATAGWATTGGAAAG 553

Qy        659 CCAATAGAGAAGGTAAATTTAT 680
              |||  || ||||: |:| | |
Db        552 GGAATTGAAAAGGWTAWTGTTT 531


RESULT 7
CNS00EVL/c
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             6.9%;  Score 53.8;  DB 9;  Length 1101;
  Best Local Similarity   34.1%;  Pred. No. 0.035;
  Matches  166;  Conservative   76;  Mismatches  243;  Indels    2;  Gaps    1;

Qy        290 TTAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATG 349
              ::|||:|  :: | : :| ||  |  ||  |: :::|:  |    |  |||||    :
Db        971 WWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTTTTTWTTTWT 912

Qy        350 TATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTT 409
                : : || |  :: :: || : :  |   :|||| | || |:|  : || | : | ||
Db        911 ATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATATTAWATTTT 852

Qy        410 AAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTT 469
                | | |  :   ||:    |||  : :|| :  : | :||    :  :     |:
Db        851 TTWTWTATTWWTATATWATWWTATWTWTWTTTWATWTATWTATTTATWTWTWTTWTATTW 792

Qy        470 CATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTT 529
              ||: || ::: : ||  : |:|   |  ||  :  | |  ::|:| |:  : ||| ||:
Db        791 TATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTWTTTTATTW 732

Qy        530 TATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTG-- 587
              | | ||:| ||:| |:|    |: |       |   | ||:    | ||    |||
Db        731 TTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTTWTATTTWAAATTATTTAAT 672

Qy        588 TGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATT 647
              | :  |:| ||| ||    |  |  ||| ||| |  |    ||| | |  ||
Db        671 TAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAATATTTT 612

Qy        648 CAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAG 707
                |   |||  :||  | |:  |||  | ::||    || :  | |:|   :   ||
Db        611 TTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTWWTTTAA 552

Qy        708 GCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTT 767
              | | |   | | ||   : : | || :  |   ::::   || :: :| | | |: ||
Db        551 ATTATWTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTWTTTTTWATT 492
```

```
Qy        768 TATTGTT 774
              :||  :|
Db        491 WATAAWT 485


RESULT 8
CNS00LT2
LOCUS       CNS00LT2                1101 bp    DNA     linear   GSS 14-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL078714
VERSION     AL078714.1  GI:5102004
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR48P19"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             6.9%;  Score 53.8;  DB 9;  Length 1101;
  Best Local Similarity   19.7%;  Pred. No. 0.035;
  Matches   91;  Conservative  163;  Mismatches  209;  Indels    0;  Gaps    0;

Qy        312 AATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTTCAAACATTAACT 371
              ||: ||   :: |||| || :   |  :|| ::  |:::: ::  |||   :  || : |
Db        569 AAKTAAWAAWWTTTTTTAATWTTTTKTWTTAKKTWAWKWWAWDTATTTWTTWTTTTTWTT 628

Qy        372 TTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGT 431
              || :  || |:  |   : |     | |  .:|:| ||   | | |  | ||| | | |
Db        629 TTTWAATTTTTWTTTWTAWATTTTTTTTTTTAKTKTTTTTTTTTTTATTAAATAAAAAAWTT 688

Qy        432 ATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAAT 491
              :   |::  |:::   |||:::::  ::| :  |  :: ::: :   :::::
Db        689 TDTWAAAWWTTTKKKKKKAAAADKWKDAKKWDGAKKATKTKKKDKKAAWAAADKKDRKKG 748

Qy        492 AGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGAT 551
              ::   |: |::|  ::: :::    :| :|:::  :::|::   :  ::  |    : :
Db        749 GKKKGKGKKGKKGGKKKKKKKKGKGGWGKKGKKKAGDDDAKDKTKKKKKAATTTTKKKK 808

Qy        552 TGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGA 611
              ::   :  :: | ::|:: : :  ||  | ::|:::|: :::| ::|  :::: : :

Db        809 GKKKKKKAAKKKKAADRTKKTKWDAAAAAAAAKKTKDKGKKKKKTKKKTTKKKKKKKKK 868

Qy        612 ATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGG 671
              : :|   :::: : :  ::|   |      :   :::: |||:    : : |  ||
Db        869 GGGKKDAAAKKKKKGTKKKKGKKGKGKGGGKGKGKGGKKDDDAAAAKKKGKKKKKGKGGG 928

Qy        672 TAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCT 731
              :   : :|:|  ::|  ::::::  :   |:| |::: | :    :: : ||: :
Db        929 GKKKKGGKKGKGKGGKKKTKKTKKKKKKKKKADAAKGKTKKRAAADAAAADWTDAATKKK 988

Qy        732 GAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTT 774
              :   ::::|:| : :  ||| : :::: : |::::| :::::
Db        989 KKKKKDKKKTKTKKKDAAAAATKKKKKKKKKTKKKKTKKKKKK 1031


RESULT 9
CNS016LI
LOCUS       CNS016LI                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106896
VERSION     AL106896.1  GI:5624374
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN16D22"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             6.9%;  Score 53.8;  DB 9;  Length 1101;
  Best Local Similarity   30.2%;  Pred. No. 0.035;
  Matches  114;  Conservative   86;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        297 CCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAAT 356
              || ||::    :| ||  || |:||:||||       |  |||||  :  |   :   |
Db        725 CCCTGWRAGGGWWAWAAAAAATTKATDTTTTTTTTTTWTTTWTTTTTTBTTTTATWTTTT 784

Qy        357 TTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTG 416
              ::      ||:| || :   ||: |: |  ||||:|       ||||: :||     |
Db        785 AKDTTTTTTTTWATATTTWTTWTTDATWTTTTTTTTTWAWTATATATTTTWAWTATAATTA 844

Qy        417 TGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCAT 476
              :    : ||  :  ||:  | :|: :: |:  ::: :: ::  |   :    | ||
Db        845 WAATWWTATTAWTATAWWATAWTWWWWWTRTAWWWTDWTWWWWATTAKTTWATWATATTA 904

Qy        477 GTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAA 536
              |:|   :|  |:: :    | :   :    |:: |||| || :| |:| |: ||::| |
Db        905 WTWAWAWWTAWAWWAWAATTATWTWTWATTATDWRTTTASTATWTTTDTWTWATAWRAKA 964
```

```
Qy        537 TAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTT 596
              ::: |:::|:|::|| :     :  | || ||  :: ::: ||| :::   | : ||
Db        965 WRWATWDWAWAWRATAWDTATWAATADATTWATATAAAWWTWWDTATTGWRWAWTADATT 1024

Qy        597 AGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAA 656
              | : | | | |    :| ::  :  || ::   : |: :| || |:: : | | ::: ::
Db       1025 AWDAAATRTAAATWTWTTKWTTKWKTTTKWTTTKTAKTKAKAGRAWWWWRTKAARRWWRW 1084

Qy        657 ATCCAATAGAGAAGGTA 673
                :||| | |   ||
Db       1085 TRTADATATATATAATA 1101


RESULT 10
CNS016CO/c
LOCUS       CNS016CO               1200 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15E04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106578
VERSION     AL106578.1  GI:5622626
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1200)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1200
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15E04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             6.7%;  Score 52.4;  DB 9;  Length 1200;
  Best Local Similarity   34.6%;  Pred. No. 0.078;
  Matches  121;  Conservative   58;  Mismatches  171;  Indels    0;  Gaps    0;

Qy        291 TAATAACCATGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGT 350
              |::|||  :  :   |  || ||  :  : ::|   :|| :   :|  |||||   || |
Db       1175 TWWTAATWWAWWTATAAATATAAATWTWWAWWAAAAWAATWTATWTAWTTTTTAWAATAT 1116

Qy        351 ATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTA 410
              :|: |:|:| : : ||||| :|:     || |:::|  |:||    : :| :|:|| ||
Db       1115 WTWTAWTWTWWTWTATTAAWWTWTTTTTTTTTWWWTTTWTTTTTTWTWATWTWTTTTTT 1056

Qy        411 AGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTC 470
              : | : :   |: :: |  ||| | ||   :   |  | : | :|   |:      :|
Db       1055 WTTWWAWTWTWTWTWWAATWTATTTTTATATTWAATAWTTTWTATWTWTTAWTWTATWTW 996

Qy        471 ATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTT 530
              |   |   :  |  :    | |   |:  :    |   |||: || | |: ||| || |
Db        995 AAAAAAAAWTTTTTWWTTTAAAATTAWAARAWAWATTAAATTWAAAAAWAWTTTTWTTRT 936

Qy        531 ATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGG 590
              : :||| | ||:| |||  || ||          |:  :|||| | :   |  | |
Db        935 TWWWAATTAATTWTTAAAATTTTTCGTTAWAATTAWTCHTTATTTTTWTARAAAATTTTT 876

Qy        591 TCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGC 640
              | ||  ||  ||:|  : | :: :|| |:    |   | : | ||:||
Db        875 AATWTTTTTTTWTTWTTWWTATWWWWAACCDAAAAAACTTTWAGHAARGC 826


RESULT 11
CNS021G2
LOCUS       CNS021G2               1146 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            225O04 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL176843
VERSION     AL176843.1  GI:7814900
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1146)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1146
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="225O04"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG225BH02LP1~end : T7"
ORIGIN

  Query Match             6.7%;  Score 52;  DB 9;  Length 1146;
  Best Local Similarity   25.6%;  Pred. No. 0.097;
  Matches   83;  Conservative  106;  Mismatches  131;  Indels    4;  Gaps    1;

Qy        452 AAAGATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGAT 511
              :||:|:: ||::: : :| |||    :||  : |    : ::  ||: :::::| :|: :
Db        561 RAADAKRAAARKKTRWDTRATTTTRADTAAAKTTAATTRRRKTTAAKRRRKKRAKKTRKK 620

Qy        512 TTTAATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTT 571
              | :: | ::: :|    ::  :::    ||:|| || :::: :   :   :   | | |
Db        621 TAKRKTTKDWAKTAARRKKTGKRRKRKKTTKTTTAAGKRKKTKTTTKTANKTAKTAAATA 680
```

```
Qy        572 ATGATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCATTGTAGGCTC 631
Db        681 ARKTTTARTTTTTDTTTAAKKRKKKRKKWAGKGKKRAKKRAKAAKTAKRRTTTTKAKATK 740

Qy        632 AGGGAAAGCAAACATTCAGAATGAAATCCAATAGAGAAGGTAAATTTATTTGGGCATGTA 691
Db        741 AARAARAAAAAAAATT----ATTKRRKBKRRKKKRRKKKRAAAAKKKBRKKTGRKGRRKG 796

Qy        692 CATTTTGGCAGCCTAGGCTGTGTACATGTGTACACATTCTGAACATGTGTGTATATTGAA 751
Db        797 GRARKKGRGAARGAAGRRKKNKKGDAWATTTTDATTATTTAATTTRADATTTTADARKK 856

Qy        752 AATCTTGTCTCTTTTTTATTGTTA 775
Db        857 KKKKKKKKKKKKTTTTTTTTTTR 880
```

RESULT 12
AG382019

```
LOCUS       AG382019               1530 bp    DNA     linear   GSS 03-JUN-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-192E18.T7, genomic survey
            sequence.
ACCESSION   AG382019
VERSION     AG382019.1  GI:47993224
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     BAC end Sequences of Library MSMg01
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1530)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1530
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-192E18.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

```
  Query Match             6.7%;  Score 51.8;  DB 9;  Length 1530;
  Best Local Similarity   49.1%;  Pred. No. 0.11;
  Matches  137;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy        335 ATCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAA 394
Db        978 ATAGTTTTTTTTAAATTTATATATGTATTTATGTTTTATTGTTATTTTTTATATTTTTAT 1037

Qy        395 GGTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAA 454
Db       1038 TTTATAGTTTTTATTTATTGTATATATTATTTATTTGTTATATATTTATTTTATTTATAG 1097

Qy        455 GATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTT 514
Db       1098 TTATTTTATTTGTTTATTTTATGGTATTTATTATTTTTTGTATTTATTATATTTTAATTT 1157

Qy        515 AATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATG 574
Db       1158 TTTATATATTATTTTTTATATATATAGTATTATTATTTATATATTATTATATGATTTATT 1217

Qy        575 ATGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAAT 613
Db       1218 TTTATAGATATTTTTATATATTTATTATATTTTTATTAT 1256
```

RESULT 13
AU088479/c

```
LOCUS       AU088479                500 bp    mRNA    linear   EST 27-JAN-2001
DEFINITION  AU088479 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA
            clone XPFn7173, mRNA sequence.
ACCESSION   AU088479
VERSION     AU088479.1  GI:12390620
KEYWORDS    EST.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
  TITLE     FULL-malaria: a database for a full-length enriched cDNA library
            from human malaria parasite, Plasmodium falciparum
  JOURNAL   Nucleic Acids Res. 29 (1), 70-71 (2001)
  MEDLINE   20574754
   PUBMED   11125052
COMMENT     Contact: Junichi Watanabe
            Institute of Medical Science
            The University of Tokyo, Department of Parasitology
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5378
            Fax: 81-3-5449-5410
            Email: jwatanab@manage.ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and
            Sugano,S. Construction and characterization of a full
            length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2),
            149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="mRNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /clone="XPFn7173"
                     /dev_stage="erythrocytic stage"
                     /clone_lib="Sugano Malaria cDNA library"
ORIGIN
```

```
  Query Match             6.6%;  Score 51.6;  DB 1;  Length 500;
  Best Local Similarity   48.1%;  Pred. No. 0.1;
  Matches  138;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy        336 TCGTTTTTGCCATGTATAAATTTTCAAACATTAACTTTCAGGGTTATTAATCCTTTTAAG 395
Db        420 TCTTTTTTTACAACTTTATAAACCTAATCAAAAATGTGTGTGTTTAATATATATATAATA 361

Qy        396 GTCTAGTTTTTCTTAAGTCTGTGCAGTAATAGAGGTATCGTCATTCATGTGACATAAAAG 455
Db        360 TTATAGGNTTATAAAAATTATGAAAATTAAAGTGTTATANATATATATATAATATTATAG 301

Qy        456 ATGGAAAGGGGCTTCATTCATGTTAGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTA 515
```

```
Db        300 GTTTATAAAAATTATGNNAATTAAAGTGTTATANATATATATATATATATATATATATAT 241

Qy        516 ATAGATGTTTCTTTTATGAAATAATTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGA 575
              ||| || | | | |||| ||||| || |||| ||||       ||   | |||||
Db        240 ATATATATATATATTATATAATAATATTAAAAATATTGGTTTATTCTATTTTATTTAATG 181

Qy        576 TGAATCATTTTGTGGTCTGTTAGTTACTTTTAGAGAATAGAAAGCAT 622
              | |||    |  |  | |  |  ||  |||   | |  | ||   ||
Db        180 TTAATTTAATATTTATTTCCTTATTTTTTTATTATATAAAAATTGAT 134


RESULT 14
CNS016BY
LOCUS       CNS016BY               1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C08 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106552
VERSION     AL106552.1  GI:5622489
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15C08"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             6.6%;  Score 51.6;  DB 9;  Length 1201;
  Best Local Similarity   29.5%;  Pred. No. 0.12;
  Matches  122;  Conservative   92;  Mismatches  199;  Indels    0;  Gaps    0;

Qy        362 AACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGCAG 421
              || | |:  :   || ||| :|  |   :: |:     | : :   : :| | | | |
Db        702 AAGAATRGGDGMRAGTGTTTBTKTTTTGWWAAMAWAAAAAAWADARWWADATTTTWTTTAW 761

Qy        422 TAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTTAG 481
              ::::||   | |  |  ||  |   | | ::|| :   |||         | |   :   |
Db        762 WWWYAGNTWTTTTTTTTTTTTTTAAAAAAWWAAAWAAAAAAAAAMAAAGRACTTGRRGGGG 821

Qy        482 TGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAATT 541
              ||   | ::: |:||::::||  || |:::|  |:|| | |     :: :|||||  :
Db        822 TGGRAGGRWDGGRAARDDRGGRAAAWTKDWTGGAWTATAAGRAAAAKKGRTGAAATTGDG 881

Qy        542 TTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGTTA 601
                |:  ||: ::  ::    :  :|  : : | | : |  ||::         ||  |:
Db        882 GAAADGGGAWGKDAARRAAGRDTTKGKGKGRGGTTTDTTTTTTWWAAAAAARWTTTTTWT 941

Qy        602 CTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATCCA 661
              :::: |:   | :   |  ||:|| :   :::    | :|:  ||  : :| |||
Db        942 TKKKKGGRATTTTKKKGGGTTTKTAADGGGRRKTTKGGGRARATTTTKKGWTAAAARRAG 1001

Qy        662 ATAGAGAAGGTAAATTTATTTGGGCATGTACATTTTGGCAGCCTAGGCTGTGTACATGTG 721
              |   : ::: ||:||| || |:: |:  |  :::|    |  : || ||:::  ::
Db       1002 TTTTTKKDKRAAARTTTTTTGGRDGAKARAATKKKTTTGTGTAKTGGGTGKKKRKGKDAA 1061

Qy        722 TACACATTCTGAACATGTGTGTATATTGAAAATCTTGTCTCTTTTTTATTGTT 774
              |   | : :|     :: ::  : : :|:: | || | | |||||| || ||
Db       1062 AAAGTAGWTWGKTTKGKKTKKGTWKKARARRTTTTTTTTTTTTTTTTTTTTT 1114


RESULT 15
CNS04D0K/c
LOCUS       CNS04D0K               945 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            101H21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL285149
VERSION     AL285149.1  GI:8023560
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 945)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="101H21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG101CD11LP1~end : T7"
ORIGIN

  Query Match             6.6%;  Score 51.4;  DB 9;  Length 945;
  Best Local Similarity   39.5%;  Pred. No. 0.13;
  Matches  145;  Conservative   40;  Mismatches  180;  Indels    2;  Gaps    1;

Qy        300 TGTGTCATCTAGAATGAAGTTATATTTTAAAAAGGATCGTTTTTGCCATGTATAAATTTT 359
              | | | ||  : |||    ||||  | |::| :    |  | |||    | |||   :| |
Db        801 TTTATTATTAWAAATTTTTTTATTTATWWATWTATTTTATATTTAATTTATATTTTWTAT 742

Qy        360 CAAACATTAACTTTCAGGGTTATTAATCCTTTTAAGGTCTAGTTTTTCTTAAGTCTGTGC 419
              | | |||   |||     |  : ::|  ::|| :  | || ||| | || | | | |
```

```
Db        741 TATATATTTTTTTTTNTWTTATWAWWTTTWWTTTWTTTTTATTTTATTTTTATTWTWTTT 682

Qy        420 AGTAATAGAGGTATCGTCATTCATGTGACATAAAAGATGGAAAGGGGCTTCATTCATGTT 479
              |  |  |  | :  : :|| :| : : :|: |   :  : |      |  ||  | ||
Db        681 TATTTTTTATATTWTAWWWTTWWTAWWWWWTWTATWTWTTWTATAAATATATTTATTTTT 622

Qy        480 AGTGATGGAAATAGGAAAGTAGGTGAAGTGATTTTAATAGATGTTTCTTTTATGAAATAA 539
              | :  |   : ||  ||| ||  |    | |   |  | || ||  |:|||   || ||
Db        621 ATWTTTATTWTTA--AAAATATTTWTTTTTAAAATTTTTTATATTATTWTTAAATAAAAA 564

Qy        540 TTTTTAAAAGATTGTCCAGCCCTGCATGATTTATGATGAATCATTTTGTGGTCTGTTAGT 599
              | |:|||| ::| |  |    :  || ||:| |  | ||  |:::| |  | | ||| |
Db        563 ATATWAAAATWWTWTAWAWAATWTTATTATWTTTTTTTAAAAAWWWTWTTATTTTTTATT 504

Qy        600 TACTTTTAGAGAATAGAAAGCATTGTAGGCTCAGGGAAAGCAAACATTCAGAATGAAATC 659
              |: :|||  | |||:      :|| ||          |||  ||: |::   |:   | |
Db        503 TWTWTTTTWAWAATWNTTTAAWTTTTATTAATTATAAAAAAAAWAAWWWTAAWRNTANTA 444

Qy        660 CAATAGA 666
              || | |
Db        443 AAAAAAA 437
```

Search completed: November 9, 2005, 04:17:53
Job time : 4017.41 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2005, 19:52:53 ; Search time 656.05 Seconds
(without alignments)
9232.381 Million cell updates/sec

Title: US-09-463-542-1_COPY_1_125
Perfect score: 125
Sequence: 1 cccctgcccctgcccctgcc.....gcgggcaggcggggcccagc 125

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	125	100.0	308	9	AF012873	AF012873 Homo sapi
	2	118.6	94.9	135675	9	AC093174	AC093174 Homo sapi
	3	118.6	94.9	180963	9	AC091492	AC091492 Homo sapi
	4	117	93.6	148828	9	AY157024	AY157024 Homo sapi
	5	62.2	49.8	108399	2	AL356736	AL356736 Homo sapi
c	6	61	48.8	238794	2	BX936377	BX936377 Danio rer
	7	59.6	47.7	171574	2	AC012300	AC012300 Homo sapi
c	8	59.4	47.5	135119	2	AC011578	AC011578 Homo sapi
	9	58.6	46.9	53121	2	AC136331	AC136331 Homo sapi
	10	58.6	46.9	68330	2	AC116109	AC116109 Mus muscu
	11	58.2	46.6	72645	2	AC112672	AC112672 Mus muscu
	12	58.2	46.6	80542	2	AC020683	AC020683 Homo sapi
	13	58	46.4	84701	2	AC016122	AC016122 Homo sapi
c	14	57.8	46.2	1416	6	AX840138	AX840138 Sequence
c	15	57.8	46.2	1416	6	AX840139	AX840139 Sequence
c	16	57.8	46.2	234009	2	CR383672	CR383672 Danio rer
	17	57.6	46.1	75449	2	AC016188	AC016188 Homo sapi
	18	57.6	46.1	79023	2	AC021951	AC021951 Homo sapi
	19	57.4	45.9	59727	2	AC100218	AC100218 Mus muscu
c	20	57.4	45.9	212028	2	CR751564	CR751564 Danio rer
	21	57.4	45.9	218236	2	AC117823	AC117823 Mus muscu
	22	57.4	45.9	240965	2	AC126054	AC126054 Mus muscu
	23	57.2	45.8	87440	2	AC021226	AC021226 Homo sapi
	24	57.2	45.8	218311	2	CR354540	CR354540 Danio rer
c	25	56.8	45.4	1060	11	BV111230	BV111230 PZA00983
	26	56.8	45.4	126667	2	AC015839	AC015839 Homo sapi
	27	56.6	45.3	66680	2	AC100483	AC100483 Mus muscu
	28	56.6	45.3	72847	2	AC023563	AC023563 Homo sapi
	29	56.6	45.3	87283	2	AC022856	AC022856 Homo sapi
c	30	56.6	45.3	180652	2	CR762436	CR762436 Danio rer
	31	56.4	45.1	78700	2	AC022685	AC022685 Homo sapi
	32	56.4	45.1	110000	2	AL929091_06	Continuation (7 of
c	33	56.4	45.1	130244	2	AC013518	AC013518 Homo sapi
c	34	56.4	45.1	186731	2	AC073212	AC073212 Homo sapi
	35	56.4	45.1	194034	9	AL592148	AL592148 Human DNA
	36	56.2	45.0	1703	9	HSHBF2A	X74143 H.sapiens H
	37	56.2	45.0	2904	9	HSHBFG2	X78202 H.sapiens H
	38	56.2	45.0	82815	2	AC062001	AC062001 Homo sapi
	39	56	44.8	51274	2	AC023454	AC023454 Homo sapi
	40	56	44.8	229518	2	AC069058	AC069058 Homo sapi
c	41	55.8	44.6	3281	2	AC143602	AC143602 Macaca mu
c	42	55.8	44.6	66596	2	AC131276	AC131276 Homo sapi
c	43	55.8	44.6	78220	2	AC023212	AC023212 Homo sapi
	44	55.8	44.6	81767	2	AC021929	AC021929 Homo sapi
	45	55.8	44.6	132745	2	AC134836	AC134836 Mus muscu

ALIGNMENTS

RESULT 1
AF012873
LOCUS AF012873 308 bp DNA linear PRI 03-MAY-1999
DEFINITION Homo sapiens PPARG gene, isoform 1, promoter and 5'UTR sequence.
ACCESSION AF012873
VERSION AF012873.1 GI:4731427
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 308)
AUTHORS Fajas,L., Auboeuf,D., Raspe,E., Schoonjans,K., Lefebvre,A.-M., Saladin,R., Najib,J., Laville,M., Fruchart,J.-C., Deeb,S., Puig-Vidal,A., Flier,J., Briggs,M., Vidal,H., Staels,B. and Auwerx,J.
TITLE The organization, promoter analysis, and expression of the human PPARgamma gene
JOURNAL J. Biol. Chem. 272 (30), 18779-18789 (1997)
MEDLINE 97373577
PUBMED 9228052
REFERENCE 2 (bases 1 to 308)
AUTHORS Fajas,L., Auwerx,J., Saladin,R. and Briggs,M.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1997) Cardiovascular Res., Ligand Pharmaceuticals, 9393 Towne Centre Dr., San Diego, CA 92121-3016, USA
FEATURES Location/Qualifiers
source 1. .308
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
gene 1. .>308
/gene="PPARG"
/note="peroxisome proliferative activated receptor, gamma"
promoter 1. .126
/gene="PPARG"
5'UTR 127. .308
/gene="PPARG"
/note="isoform 1"
ORIGIN

```
  Query Match           100.0%;  Score 125;  DB 9;  Length 308;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-09;
  Matches  125;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60

Qy         61 GCCCGCCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCCGCCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120

Qy        121 CCAGC 125
              |||||
Db        121 CCAGC 125
```

RESULT 2
AC093174
LOCUS AC093174 135675 bp DNA linear PRI 08-NOV-2002
DEFINITION Homo sapiens chromosome 3 clone RP11-167M22 map 3p, complete sequence.
ACCESSION AC093174
VERSION AC093174.2 GI:24796717
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 135675)
AUTHORS Wu,Q., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L., Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y., Wu,D., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and Yang,H.
TITLE Chromosome 3p genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 135675)
AUTHORS Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L., Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and Yang,H.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-2001) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China
REFERENCE 3 (bases 1 to 135675)
AUTHORS Wu,Q., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L., Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y., Wu,D., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and Yang,H.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2002) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China
COMMENT On Nov 8, 2002 this sequence version replaced gi:15148930.
-------------Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgc.igtp.ac.cn
http://www.genomics.org.cn
Contact:hgc@igtp.ac.cn
-------------- Project Information
Center project name:1% project
Center clone name: RP11-167M22
-------------- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator: ET 55% of reads
Chemistry: Dye-terminator Big Dye; 45% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 0 bases at least Q40
Consensus quality: 0 bases at least Q30
Consensus quality: 6 bases at least Q20
Insert size: 3392; sum-of-contigs
Quality coverage: 0.00x in Q20 bases;sum-of-contigs
--------------.
FEATURES Location/Qualifiers
source 1. .135675
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-167M22"
ORIGIN

```
  Query Match           94.9%;  Score 118.6;  DB 9;  Length 135675;
  Best Local Similarity  96.8%;  Pred. No. 5.7e-09;
  Matches  121;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      18026 CCCCCGCCCCCGCCCCCGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 18085

Qy         61 GCCCGCCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      18086 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGCTCCGCCGCGGGCAGGCGGGGC 18145

Qy        121 CCAGC 125
              |||||
Db      18146 CCAGC 18150
```

RESULT 3
AC091492
LOCUS AC091492 180963 bp DNA linear PRI 20-DEC-2002
DEFINITION Homo sapiens chromosome 3 clone RP11-335I9 map 3p, complete sequence.
ACCESSION AC091492
VERSION AC091492.2 GI:24796732
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 180963)
AUTHORS Bao,W., Bao,J., Bao,Q., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L., Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,

```
            Yu,J. and Yang,H.
  TITLE     Chromosome 3p genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180963)
  AUTHORS   Bao,W., Bao,J., Bao,Q., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2001) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
REFERENCE   3  (bases 1 to 180963)
  AUTHORS   Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
REFERENCE   4  (bases 1 to 180963)
  AUTHORS   Bao,W., Bao,J., Bao,Q., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,W., Li,Y., Luo,J., Niu,Y., Qi,Q., Qi,x., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y.,
            Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B.,
            Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-2002) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
COMMENT     On Nov 8, 2002 this sequence version replaced gi:13794231.
            --------------Genome Center
            Center:Beijing Center
            Center code:Beijing
            Website:http://hgc.igtp.ac.cn
            http://www.genomics.org.cn
            Contact:hgc@igtp.ac.cn
            -------------- Project Information
            Center project name:1% project
            Center clone name: RP11-335I9
            -------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator: ET 55% of reads
            Chemistry: Dye-terminator Big Dye; 45% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 369 bases at least Q40
            Consensus quality: 571 bases at least Q30
            Consensus quality: 755 bases at least Q20
            Insert size: 773; sum-of-contigs
            Quality coverage:   1.48x in Q20 bases;sum-of-contigs
            --------------.
FEATURES              Location/Qualifiers
     source           1. .180963
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="3"
                      /map="3p"
                      /clone="RP11-335I9"
ORIGIN

  Query Match             94.9%;  Score 118.6;  DB 9;  Length 180963;
  Best Local Similarity   96.8%;  Pred. No. 5.1e-09;
  Matches  121;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     138044 CCCCCGCCCCCGCCCCCGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 138103

Qy         61 GCCCGCCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     138104 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGCTCCGCCGCGGGCAGGCGGGGC 138163

Qy        121 CCAGC 125
              |||||
Db     138164 CCAGC 138168


RESULT 4
AY157024
LOCUS       AY157024               148828 bp    DNA     linear   PRI 16-JUN-2004
DEFINITION  Homo sapiens peroxisome proliferative activated receptor gamma
            (PPARG) gene, complete cds.
ACCESSION   AY157024
VERSION     AY157024.2  GI:48762804
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148828)
  AUTHORS   Rieder,M.J., da Ponte,S.H., Kuldanek,S.A., Rajkumar,N., Smith,J.D.,
            Toth,E.J. and Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2002) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
REFERENCE   2  (bases 1 to 148828)
  AUTHORS   Rieder,M.J., Daniels,R.L., da Ponte,S.H., Hastings,N.C.,
            Ahearn,M.O., Rajkumar,N., Yi,Q. and Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-2004) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
  REMARK    Sequence update by submitter
COMMENT     On Jun 16, 2004 this sequence version replaced gi:23953882.
            To cite this work please use:  SeattleSNPs. NHLBI Program for
            Genomic Applications, UW-FHCRC, Seattle, WA
            (URL:http://pga.mbt.washington.edu).
FEATURES              Location/Qualifiers
     source           1. .148828
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
     repeat_region    9. .59
                      /rpt_family="L1"
                      /rpt_type=dispersed
     variation        54
                      /frequency="0.12"
                      /replace="g"
     repeat_region    60. .351
                      /rpt_family="Alu"
                      /rpt_type=dispersed
     variation        147
                      /frequency="0.03"
                      /replace="t"
     variation        252
```

```
                /frequency="0.03"
                /replace="g"
repeat_region   352. .431
                /rpt_family="L1"
                /rpt_type=dispersed
variation       374
                /frequency="0.05"
                /replace="c"
repeat_region   435. .759
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1555
                /frequency="0.02"
                /replace="t"
misc_feature    1754. .1925
                /note="Region not scanned for variation"
repeat_region   1784. .2127
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   2032. .2632
                /rpt_family="L1"
                /rpt_type=dispersed
gene            2048. .148443
                /gene="PPARG"
mRNA            join(2048. .2143,26500. .26573,93809. .94036,95427. .95596,
                106715. .106853,119980. .120179,130803. .131253,
                147997. .148443)
                /gene="PPARG"
                /product="peroxisome proliferative activated receptor
                gamma"
variation       2076
                /gene="PPARG"
                /frequency="0.15"
                /replace="t"
variation       2107
                /gene="PPARG"
                /frequency="0.01"
                /replace="a"
variation       2186
                /gene="PPARG"
                /frequency="0.22"
                /replace="c"
variation       2382
                /gene="PPARG"
                /frequency="0.01"
                /replace="t"
variation       2407
                /gene="PPARG"
                /frequency="0.05"
                /replace="t"
variation       2499
                /gene="PPARG"
                /frequency="0.04"
                /replace="t"
variation       2767
                /gene="PPARG"
                /frequency="0.07"
                /replace="g"
variation       3035
                /gene="PPARG"
                /frequency="0.01"
                /replace="a"
variation       3198
                /gene="PPARG"
                /frequency="0.04"
                /replace="t"
variation       3354
                /gene="PPARG"
                /frequency="0.01"
                /replace=""
variation       4160
                /gene="PPARG"
                /frequency="0.01"
                /replace="c"
misc_feature    4220. .25559
                /gene="PPARG"
                /note="Region not scanned for variation"
repeat_region   5952. .6004
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   8735. .9019
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   9637. .9935
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   10092. .10404
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   10423. .10537
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   11193. .11356
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   12140. .12449
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   12949. .13255
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   13277. .13383
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   16397. .16653
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   17063. .17196
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   17325. .17622
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   17653. .17733
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   17766. .17984
                /rpt_family="CR1"
                /rpt_type=dispersed
repeat_region   18787. .18845
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   19727. .19804
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   19757. .19813
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   20083. .20181
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   20182. .20524
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   20525. .20660
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   20937. .21239
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   21961. .22204
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   22205. .22492
                /rpt_family="Alu"
                /rpt_type=dispersed
```

```
     repeat_region   22493. .22581
                     /rpt_family="L2"
                     /rpt_type=dispersed
     repeat_region   22582. .22916
                     /rpt_family="MER1_type"
                     /rpt_type=dispersed
     repeat_region   22917. .23104
                     /rpt_family="L2"
                     /rpt_type=dispersed

  Query Match             93.6%;  Score 117;  DB 9;  Length 148828;
  Best Local Similarity   96.0%;  Pred. No. 9.2e-09;
  Matches  120;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCCAGCCGGCGCCCGC 60
              |||| ||||| ||||| |||||| ||||||||||||||||||||||||||||||||||||
Db       1922 CCCCCGCCCCCGCCCCCGCCCCCGCCCCCACCCCCACCCCCACCCCCCAGCCGGCGCCCGC 1981

Qy         61 GCCCGCCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1982 GCCCGCCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGCTCCGCCGCGGGCAGGCGGGGC 2041

Qy        121 CCAGC 125
              |||||
Db       2042 CCAGC 2046


RESULT 5
AL356736
LOCUS       AL356736              108399 bp    DNA     linear   HTG 13-JUN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-422P10, 26 unordered pieces.
ACCESSION   AL356736
VERSION     AL356736.3  GI:9926625
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 25, 2000 this sequence version replaced gi:9213961.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA422P10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 95168 bases at least Q40
            Consensus quality: 100303 bases at least Q30
            Consensus quality: 102890 bases at least Q20
            Insert size: 105899; sum-of-contigs
            Insert size: 156533; 8.5% error; agarose-fp
            Quality coverage: 2.72x in Q20 bases; sum-of-contigs Quality
            coverage: 2.16x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2189: contig of 2189 bp in length
            *     2190     2289: gap of 100 bp
            *     2290    10136: contig of 7847 bp in length
            *    10137    10236: gap of 100 bp
            *    10237    14836: contig of 4600 bp in length
            *    14837    14936: gap of 100 bp
            *    14937    20068: contig of 5132 bp in length
            *    20069    20168: gap of 100 bp
            *    20169    23196: contig of 3028 bp in length
            *    23197    23296: gap of 100 bp
            *    23297    25747: contig of 2451 bp in length
            *    25748    25847: gap of 100 bp
            *    25848    28603: contig of 2756 bp in length
            *    28604    28703: gap of 100 bp
            *    28704    33157: contig of 4454 bp in length
            *    33158    33257: gap of 100 bp
            *    33258    38474: contig of 5217 bp in length
            *    38475    38574: gap of 100 bp
            *    38575    44658: contig of 6084 bp in length
            *    44659    44758: gap of 100 bp
            *    44759    49335: contig of 4577 bp in length
            *    49336    49435: gap of 100 bp
            *    49436    51799: contig of 2364 bp in length
            *    51800    51899: gap of 100 bp
            *    51900    54349: contig of 2450 bp in length
            *    54350    54449: gap of 100 bp
            *    54450    60488: contig of 6039 bp in length
            *    60489    60588: gap of 100 bp
            *    60589    67658: contig of 7070 bp in length
            *    67659    67758: gap of 100 bp
            *    67759    69787: contig of 2029 bp in length
            *    69788    69887: gap of 100 bp
            *    69888    77142: contig of 7255 bp in length
            *    77143    77242: gap of 100 bp
            *    77243    81918: contig of 4676 bp in length
            *    81919    82018: gap of 100 bp
            *    82019    86773: contig of 4755 bp in length
            *    86774    86873: gap of 100 bp
            *    86874    88965: contig of 2092 bp in length
            *    88966    89065: gap of 100 bp
            *    89066    91615: contig of 2550 bp in length
            *    91616    91715: gap of 100 bp
            *    91716    95214: contig of 3499 bp in length
            *    95215    95314: gap of 100 bp
            *    95315    99815: contig of 4501 bp in length
            *    99816    99915: gap of 100 bp
            *    99916   101937: contig of 2022 bp in length
            *   101938   102037: gap of 100 bp
            *   102038   105570: contig of 3533 bp in length
            *   105571   105670: gap of 100 bp
            *   105671   108399: contig of 2729 bp in length.
FEATURES             Location/Qualifiers
     source          1. .108399
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-422P10"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .2189
                     /note="assembly_fragment:00802
                     fragment_chain:1"
     misc_feature    2290. .10136
                     /note="assembly_fragment:01048
                     fragment_chain:1"
     misc_feature    10237. .14836
                     /note="assembly_fragment:00280
                     fragment_chain:1"
     misc_feature    14937. .20068
                     /note="assembly_fragment:00172
                     fragment_chain:1"
     misc_feature    20169. .23196
                     /note="assembly_fragment:00258
                     fragment_chain:1"
```

```
     misc_feature    23297. .25747
                     /note="assembly_fragment:00737
                     fragment_chain:1"
     misc_feature    25848. .28603
                     /note="assembly_fragment:00087
                     fragment_chain:2"
     misc_feature    28704. .33157
                     /note="assembly_fragment:00528
                     fragment_chain:2"
     misc_feature    33258. .38474
                     /note="assembly_fragment:00740
                     fragment_chain:2"
     misc_feature    38575. .44658
                     /note="assembly_fragment:00210
                     fragment_chain:3"
     misc_feature    44759. .49335
                     /note="assembly_fragment:00597
                     fragment_chain:3"
     misc_feature    49436. .51799
                     /note="assembly_fragment:01053
                     fragment_chain:3"
     misc_feature    51900. .54349
                     /note="assembly_fragment:00005
                     fragment_chain:4"
     misc_feature    54450. .60488
                     /note="assembly_fragment:00750
                     fragment_chain:4"
     misc_feature    60589. .67658
                     /note="assembly_fragment:00371
                     fragment_chain:5"
     misc_feature    67759. .69787
                     /note="assembly_fragment:01118
                     fragment_chain:5"
     misc_feature    69888. .77142
                     /note="assembly_fragment:00679
                     fragment_chain:6"
     misc_feature    77243. .81918
                     /note="assembly_fragment:00766
                     fragment_chain:6"
     misc_feature    82019. .86773
                     /note="assembly_fragment:00459"
     misc_feature    86874. .88965
                     /note="assembly_fragment:00571"
     misc_feature    89066. .91615
                     /note="assembly_fragment:00655"
     misc_feature    91716. .95214
                     /note="assembly_fragment:00708"
     misc_feature    95315. .99815
                     /note="assembly_fragment:00875"
     misc_feature    99916. .101937
                     /note="assembly_fragment:00967"
     misc_feature    102038. .105570
                     /note="assembly_fragment:01044"
     misc_feature    105671. .108399
                     /note="assembly_fragment:00130
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             49.8%;  Score 62.2;  DB 2;  Length 108399;
  Best Local Similarity   68.0%;  Pred. No. 0.61;
  Matches   85;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| |||| |||| ||||| ||||| ||||| ||||| ||||| ||  | ||| |
Db       2062 CCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 2121

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              ||||||||| | ||   |||  | |  |||  ||||   || |||||  | | |||||
Db       2122 CCCCGCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCGCCCCCCCCGCGCCCCGCCGGGGG 2181

Qy        121 CCAGC 125
                | ||
```

```
Db       2182 GCGGC 2186


RESULT 6
BX936377/c
LOCUS       BX936377              238794 bp    DNA     linear   HTG 10-FEB-2004
DEFINITION  Danio rerio clone DKEY-149J2, *** SEQUENCING IN PROGRESS ***, 14
            unordered pieces.
ACCESSION   BX936377
VERSION     BX936377.3  GI:42516919
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 238794)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 10, 2004 this sequence version replaced gi:42491169.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK149J2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 231615 bases at least Q40
            Consensus quality: 232776 bases at least Q30
            Consensus quality: 234285 bases at least Q20
            Insert size: 237494; sum-of-contigs
            Insert size: 196119; 7.4% error; agarose-fp
            Quality coverage: 8.08x in Q20 bases; sum-of-contigs Quality
            coverage: 9.83x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    30643: contig of 30643 bp in length
            *    30644    30743: gap of 100 bp
            *    30744    33315: contig of 2572 bp in length
            *    33316    33415: gap of 100 bp
            *    33416    48607: contig of 15192 bp in length
            *    48608    48707: gap of 100 bp
            *    48708    66118: contig of 17411 bp in length
            *    66119    66218: gap of 100 bp
            *    66219    69115: contig of 2897 bp in length
            *    69116    69215: gap of 100 bp
            *    69216    81253: contig of 12038 bp in length
            *    81254    81353: gap of 100 bp
            *    81354   114907: contig of 33554 bp in length
            *   114908   115007: gap of 100 bp
            *   115008   160636: contig of 45629 bp in length
            *   160637   160736: gap of 100 bp
            *   160737   169443: contig of 8707 bp in length
            *   169444   169543: gap of 100 bp
            *   169544   179161: contig of 9618 bp in length
            *   179162   179261: gap of 100 bp
            *   179262   196015: contig of 16754 bp in length
            *   196016   196115: gap of 100 bp
            *   196116   230887: contig of 34772 bp in length
            *   230888   230987: gap of 100 bp
```

```
           *   230988   235549: contig of 4562 bp in length
           *   235550   235649: gap of 100 bp
           *   235650   238794: contig of 3145 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238794
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-149J2"
                     /clone_lib="DanioKey"
     misc_feature    1. .30643
                     /note="assembly_fragment:01093
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    30744. .33315
                     /note="assembly_fragment:00031
                     fragment_chain:1"
     misc_feature    33416. .48607
                     /note="assembly_fragment:00681
                     fragment_chain:2"
     misc_feature    48708. .66118
                     /note="assembly_fragment:00385
                     fragment_chain:2"
     misc_feature    66219. .69115
                     /note="assembly_fragment:00009
                     fragment_chain:2"
     misc_feature    69216. .81253
                     /note="assembly_fragment:00527
                     fragment_chain:2"
     misc_feature    81354. .114907
                     /note="assembly_fragment:01480
                     fragment_chain:3"
     misc_feature    115008. .160636
                     /note="assembly_fragment:02307
                     fragment_chain:3"
     misc_feature    160737. .169443
                     /note="assembly_fragment:00135
                     fragment_chain:3"
     misc_feature    169544. .179161
                     /note="assembly_fragment:00253
                     fragment_chain:3"
     misc_feature    179262. .196015
                     /note="assembly_fragment:00868
                     fragment_chain:3"
     misc_feature    196116. .230887
                     /note="assembly_fragment:01889
                     fragment_chain:3"
     misc_feature    230988. .235549
                     /note="assembly_fragment:00088
                     fragment_chain:3"
     misc_feature    235650. .238794
                     /note="assembly_fragment:00054
                     fragment_chain:3
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             48.8%;  Score 61;  DB 2;  Length 238794;
  Best Local Similarity   68.0%;  Pred. No. 0.67;
  Matches   85;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  ||||  ||||| ||||| ||||| ||||| ||||| ||  |||| |||
Db     169926 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCGCGCGCGCGC 169867

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||||||||| | ||   |||  | |  |||  ||| | |   || | |   |||| ||
Db     169866 GCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCGCGCGCGGGCGGGCGCGGCGC 169807

Qy        121 CCAGC 125
              |  ||
Db     169806 CGCGC 169802
```

```
RESULT 7
AC012300
LOCUS       AC012300              171574 bp    DNA     linear   HTG 12-MAR-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-248L3 map 17, *** SEQUENCING
            IN PROGRESS ***, 48 unordered pieces.
ACCESSION   AC012300
VERSION     AC012300.2  GI:7230237
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171574)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-248L3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171574)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6094565.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L670
                Center clone name: 248_L_3
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 48 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1038: contig of 1038 bp in length
            *     1039     1138: gap of 100 bp
            *     1139     2373: contig of 1235 bp in length
            *     2374     2473: gap of 100 bp
            *     2474     3537: contig of 1064 bp in length
            *     3538     3637: gap of 100 bp
            *     3638     4776: contig of 1139 bp in length
            *     4777     4876: gap of 100 bp
            *     4877     6026: contig of 1150 bp in length
            *     6027     6126: gap of 100 bp
            *     6127     7205: contig of 1079 bp in length
            *     7206     7305: gap of 100 bp
            *     7306     8344: contig of 1039 bp in length
            *     8345     8444: gap of 100 bp
            *     8445     9543: contig of 1099 bp in length
            *     9544     9643: gap of 100 bp
```

```
*     9644    10869: contig of 1226 bp in length
*    10870    10969: gap of 100 bp
*    10970    12118: contig of 1149 bp in length
*    12119    12218: gap of 100 bp
*    12219    13641: contig of 1423 bp in length
*    13642    13741: gap of 100 bp
*    13742    15584: contig of 1843 bp in length
*    15585    15684: gap of 100 bp
*    15685    17014: contig of 1330 bp in length
*    17015    17114: gap of 100 bp
*    17115    18904: contig of 1790 bp in length
*    18905    19004: gap of 100 bp
*    19005    20212: contig of 1208 bp in length
*    20213    20312: gap of 100 bp
*    20313    21330: contig of 1018 bp in length
*    21331    21430: gap of 100 bp
*    21431    22568: contig of 1138 bp in length
*    22569    22668: gap of 100 bp
*    22669    24914: contig of 2246 bp in length
*    24915    25014: gap of 100 bp
*    25015    26711: contig of 1697 bp in length
*    26712    26811: gap of 100 bp
*    26812    28657: contig of 1846 bp in length
*    28658    28757: gap of 100 bp
*    28758    30770: contig of 2013 bp in length
*    30771    30870: gap of 100 bp
*    30871    32848: contig of 1978 bp in length
*    32849    32948: gap of 100 bp
*    32949    35389: contig of 2441 bp in length
*    35390    35489: gap of 100 bp
*    35490    37527: contig of 2038 bp in length
*    37528    37627: gap of 100 bp
*    37628    40579: contig of 2952 bp in length
*    40580    40679: gap of 100 bp
*    40680    43521: contig of 2842 bp in length
*    43522    43621: gap of 100 bp
*    43622    45933: contig of 2312 bp in length
*    45934    46033: gap of 100 bp
*    46034    48737: contig of 2704 bp in length
*    48738    48837: gap of 100 bp
*    48838    51899: contig of 3062 bp in length
*    51900    51999: gap of 100 bp
*    52000    55449: contig of 3450 bp in length
*    55450    55549: gap of 100 bp
*    55550    60135: contig of 4586 bp in length
*    60136    60235: gap of 100 bp
*    60236    63645: contig of 3410 bp in length
*    63646    63745: gap of 100 bp
*    63746    66221: contig of 2476 bp in length
*    66222    66321: gap of 100 bp
*    66322    71307: contig of 4986 bp in length
*    71308    71407: gap of 100 bp
*    71408    75445: contig of 4038 bp in length
*    75446    75545: gap of 100 bp
*    75546    81775: contig of 6230 bp in length
*    81776    81875: gap of 100 bp
*    81876    87070: contig of 5195 bp in length
*    87071    87170: gap of 100 bp
*    87171    91815: contig of 4645 bp in length
*    91816    91915: gap of 100 bp
*    91916    98112: contig of 6197 bp in length
*    98113    98212: gap of 100 bp
*    98213   103688: contig of 5476 bp in length
*   103689   103788: gap of 100 bp
*   103789   111263: contig of 7475 bp in length
*   111264   111363: gap of 100 bp
*   111364   118151: contig of 6788 bp in length
*   118152   118251: gap of 100 bp
*   118252   124362: contig of 6111 bp in length
*   124363   124462: gap of 100 bp
*   124463   133078: contig of 8616 bp in length
*   133079   133178: gap of 100 bp
*   133179   139647: contig of 6469 bp in length
*   139648   139747: gap of 100 bp
*   139748   146970: contig of 7223 bp in length
*   146971   147070: gap of 100 bp
*   147071   156784: contig of 9714 bp in length
*   156785   156884: gap of 100 bp
*   156885   171574: contig of 14690 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171574
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-248L3"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1038
                     /note="assembly_fragment"
     misc_feature    1139. .2373
                     /note="assembly_fragment"
     misc_feature    2474. .3537
                     /note="assembly_fragment"
     misc_feature    3638. .4776
                     /note="assembly_fragment"
     misc_feature    4877. .6026
                     /note="assembly_fragment"
     misc_feature    6127. .7205
                     /note="assembly_fragment"
     misc_feature    7306. .8344
                     /note="assembly_fragment"
     misc_feature    8445. .9543
                     /note="assembly_fragment"
     misc_feature    9644. .10869
                     /note="assembly_fragment"
     misc_feature    10970. .12118
                     /note="assembly_fragment"
     misc_feature    12219. .13641
                     /note="assembly_fragment"
     misc_feature    13742. .15584
                     /note="assembly_fragment"
     misc_feature    15685. .17014
                     /note="assembly_fragment"
     misc_feature    17115. .18904
                     /note="assembly_fragment"
     misc_feature    19005. .20212
                     /note="assembly_fragment"
     misc_feature    20313. .21330
                     /note="assembly_fragment"
     misc_feature    21431. .22568
                     /note="assembly_fragment"
     misc_feature    22669. .24914
                     /note="assembly_fragment"
     misc_feature    25015. .26711
                     /note="assembly_fragment"
     misc_feature    26812. .28657
                     /note="assembly_fragment"
     misc_feature    28758. .30770
                     /note="assembly_fragment"
     misc_feature    30871. .32848

  Query Match             47.7%;  Score 59.6;  DB 2;  Length 171574;
  Best Local Similarity   68.0%;  Pred. No. 1.2;
  Matches   83;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| ||||| ||||||  |||| ||||| | ||| ||||| |||||  ||  | ||| |
Db      20712 CCCCCGCCCCCGCCCCTTTCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCC 20771

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||| ||||| ||||   |||| | |   ||  | |    ||||| ||    | | ||||
Db      20772 GCCCCCCCCCCCGCCCCCCCCGCCCCCCTCCCGCGCCCCCCCGCCCCGCCGCGCCCGGGC 20831

Qy        121 CC 122
              ||
```

```
Db      20832 CC 20833


RESULT 8
AC011578/c
LOCUS       AC011578              135119 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-12P9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC011578
VERSION     AC011578.4  GI:9123847
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 135119)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-12P9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 135119)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:7107923.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3313
                Center clone name: 12_P_9
            --------------.
            * NOTE: This record contains 146 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      744: contig of 744 bp in length
            *      745      844: gap of 100 bp
            *      845     1615: contig of 771 bp in length
            *     1616     1715: gap of 100 bp
            *     1716     2485: contig of 770 bp in length
            *     2486     2585: gap of 100 bp
            *     2586     3345: contig of 760 bp in length
            *     3346     3445: gap of 100 bp
            *     3446     4188: contig of 743 bp in length
            *     4189     4288: gap of 100 bp
            *     4289     5050: contig of 762 bp in length
            *     5051     5150: gap of 100 bp
            *     5151     5984: contig of 834 bp in length
            *     5985     6084: gap of 100 bp
            *     6085     6837: contig of 753 bp in length
            *     6838     6937: gap of 100 bp
            *     6938     7694: contig of 757 bp in length
            *     7695     7794: gap of 100 bp
            *     7795     8541: contig of 747 bp in length
            *     8542     8641: gap of 100 bp
            *     8642     9403: contig of 762 bp in length
            *     9404     9503: gap of 100 bp
            *     9504    10273: contig of 770 bp in length
            *    10274    10373: gap of 100 bp
            *    10374    11138: contig of 765 bp in length
            *    11139    11238: gap of 100 bp
            *    11239    11991: contig of 753 bp in length
            *    11992    12091: gap of 100 bp
            *    12092    12840: contig of 749 bp in length
            *    12841    12940: gap of 100 bp
            *    12941    13800: contig of 860 bp in length
            *    13801    13900: gap of 100 bp
            *    13901    14661: contig of 761 bp in length
            *    14662    14761: gap of 100 bp
            *    14762    15508: contig of 747 bp in length
            *    15509    15608: gap of 100 bp
            *    15609    16359: contig of 751 bp in length
            *    16360    16459: gap of 100 bp
            *    16460    17225: contig of 766 bp in length
            *    17226    17325: gap of 100 bp
            *    17326    18074: contig of 749 bp in length
            *    18075    18174: gap of 100 bp
            *    18175    18916: contig of 742 bp in length
            *    18917    19016: gap of 100 bp
            *    19017    19785: contig of 769 bp in length
            *    19786    19885: gap of 100 bp
            *    19886    20641: contig of 756 bp in length
            *    20642    20741: gap of 100 bp
            *    20742    21501: contig of 760 bp in length
            *    21502    21601: gap of 100 bp
            *    21602    22367: contig of 766 bp in length
            *    22368    22467: gap of 100 bp
            *    22468    23213: contig of 746 bp in length
            *    23214    23313: gap of 100 bp
            *    23314    24069: contig of 756 bp in length
            *    24070    24169: gap of 100 bp
            *    24170    24922: contig of 753 bp in length
            *    24923    25022: gap of 100 bp
            *    25023    25771: contig of 749 bp in length
            *    25772    25871: gap of 100 bp
            *    25872    26632: contig of 761 bp in length
            *    26633    26732: gap of 100 bp
            *    26733    27495: contig of 763 bp in length
            *    27496    27595: gap of 100 bp
            *    27596    28366: contig of 771 bp in length
            *    28367    28466: gap of 100 bp
            *    28467    29222: contig of 756 bp in length
            *    29223    29322: gap of 100 bp
            *    29323    30092: contig of 770 bp in length
            *    30093    30192: gap of 100 bp
            *    30193    30945: contig of 753 bp in length
            *    30946    31045: gap of 100 bp
            *    31046    31793: contig of 748 bp in length
            *    31794    31893: gap of 100 bp
            *    31894    32649: contig of 756 bp in length
            *    32650    32749: gap of 100 bp
            *    32750    33490: contig of 741 bp in length
            *    33491    33590: gap of 100 bp
            *    33591    34331: contig of 741 bp in length
            *    34332    34431: gap of 100 bp
            *    34432    35176: contig of 745 bp in length
            *    35177    35276: gap of 100 bp
            *    35277    36039: contig of 763 bp in length
            *    36040    36139: gap of 100 bp
            *    36140    36886: contig of 747 bp in length
            *    36887    36986: gap of 100 bp
```

```
            *    36987    37722: contig of 736 bp in length
            *    37723    37822: gap of 100 bp
            *    37823    38465: contig of 643 bp in length
            *    38466    38565: gap of 100 bp
            *    38566    39336: contig of 771 bp in length
            *    39337    39436: gap of 100 bp
            *    39437    40204: contig of 768 bp in length
            *    40205    40304: gap of 100 bp
            *    40305    41065: contig of 761 bp in length
            *    41066    41165: gap of 100 bp
            *    41166    41912: contig of 747 bp in length
            *    41913    42012: gap of 100 bp
            *    42013    42781: contig of 769 bp in length
            *    42782    42881: gap of 100 bp
            *    42882    43658: contig of 777 bp in length
            *    43659    43758: gap of 100 bp
            *    43759    44517: contig of 759 bp in length
            *    44518    44617: gap of 100 bp
            *    44618    45375: contig of 758 bp in length
            *    45376    45475: gap of 100 bp
            *    45476    46206: contig of 731 bp in length
            *    46207    46306: gap of 100 bp
            *    46307    47061: contig of 755 bp in length
            *    47062    47161: gap of 100 bp
            *    47162    47915: contig of 754 bp in length
            *    47916    48015: gap of 100 bp
            *    48016    48759: contig of 744 bp in length
            *    48760    48859: gap of 100 bp
            *    48860    49607: contig of 748 bp in length
            *    49608    49707: gap of 100 bp
            *    49708    50493: contig of 786 bp in length
            *    50494    50593: gap of 100 bp
            *    50594    51352: contig of 759 bp in length
            *    51353    51452: gap of 100 bp
            *    51453    52197: contig of 745 bp in length
            *    52198    52297: gap of 100 bp
            *    52298    53039: contig of 742 bp in length
            *    53040    53139: gap of 100 bp
            *    53140    53903: contig of 764 bp in length
            *    53904    54003: gap of 100 bp
            *    54004    54763: contig of 760 bp in length
            *    54764    54863: gap of 100 bp
            *    54864    55629: contig of 766 bp in length
            *    55630    55729: gap of 100 bp
            *    55730    56491: contig of 762 bp in length
            *    56492    56591: gap of 100 bp
            *    56592    57319: contig of 728 bp in length
            *    57320    57419: gap of 100 bp
            *    57420    58171: contig of 752 bp in length
            *    58172    58271: gap of 100 bp
            *    58272    59016: contig of 745 bp in length
            *    59017    59116: gap of 100 bp
            *    59117    59863: contig of 747 bp in length
            *    59864    59963: gap of 100 bp
            *    59964    60734: contig of 771 bp in length
            *    60735    60834: gap of 100 bp
            *    60835    61594: contig of 760 bp in length
            *    61595    61694: gap of 100 bp
            *    61695    62469: contig of 775 bp in length

  Query Match             47.5%;  Score 59.4;  DB 2;  Length 135119;
  Best Local Similarity   69.2%;  Pred. No. 1.4;
  Matches   81;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          8 CCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCCCGCC 67
              |||  |||| |||||  ||| | ||||| ||||  |||||  ||||| |||||  | |||
Db       5570 CCCACCCCCCGCCCCACCCCGCCCCCCCCCCCCCGCCCCCCCCCCGGCCCCCGCCGCAGCC 5511

Qy         68 CCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGCCCAG 124
              | | |||| ||| |||| | ||||| | |    || |  ||||  |||||||  | |
Db       5510 CGCCCGCCCGGCGCGGCGCCGCCCGGCGCCCCCCCCCGCCGGGGCGGCGGGGGGCCG 5454
```

```
RESULT 9
AC136331
LOCUS       AC136331               53121 bp    DNA     linear   HTG 14-NOV-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-1230B22 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC136331
VERSION     AC136331.2  GI:24962744
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 53121)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-1230B22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 53121)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 53121)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 14, 2002 this sequence version replaced gi:24431749.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28484
                Center clone name: 1230_B_22
            --------------.
```

```
* NOTE: This record contains 66 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      700: contig of 700 bp in length
*      701      800: gap of 100 bp
*      801     1495: contig of 695 bp in length
*     1496     1595: gap of 100 bp
*     1596     2308: contig of 713 bp in length
*     2309     2408: gap of 100 bp
*     2409     3141: contig of 733 bp in length
*     3142     3241: gap of 100 bp
*     3242     3972: contig of 731 bp in length
*     3973     4072: gap of 100 bp
*     4073     4788: contig of 716 bp in length
*     4789     4888: gap of 100 bp
*     4889     5595: contig of 707 bp in length
*     5596     5695: gap of 100 bp
*     5696     6415: contig of 720 bp in length
*     6416     6515: gap of 100 bp
*     6516     7221: contig of 706 bp in length
*     7222     7321: gap of 100 bp
*     7322     8019: contig of 698 bp in length
*     8020     8119: gap of 100 bp
*     8120     8813: contig of 694 bp in length
*     8814     8913: gap of 100 bp
*     8914     9646: contig of 733 bp in length
*     9647     9746: gap of 100 bp
*     9747    10436: contig of 690 bp in length
*    10437    10536: gap of 100 bp
*    10537    11235: contig of 699 bp in length
*    11236    11335: gap of 100 bp
*    11336    12041: contig of 706 bp in length
*    12042    12141: gap of 100 bp
*    12142    12870: contig of 729 bp in length
*    12871    12970: gap of 100 bp
*    12971    13639: contig of 669 bp in length
*    13640    13739: gap of 100 bp
*    13740    14435: contig of 696 bp in length
*    14436    14535: gap of 100 bp
*    14536    15235: contig of 700 bp in length
*    15236    15335: gap of 100 bp
*    15336    16062: contig of 727 bp in length
*    16063    16162: gap of 100 bp
*    16163    16862: contig of 700 bp in length
*    16863    16962: gap of 100 bp
*    16963    17662: contig of 700 bp in length
*    17663    17762: gap of 100 bp
*    17763    18472: contig of 710 bp in length
*    18473    18572: gap of 100 bp
*    18573    19283: contig of 711 bp in length
*    19284    19383: gap of 100 bp
*    19384    20092: contig of 709 bp in length
*    20093    20192: gap of 100 bp
*    20193    20905: contig of 713 bp in length
*    20906    21005: gap of 100 bp
*    21006    21712: contig of 707 bp in length
*    21713    21812: gap of 100 bp
*    21813    22503: contig of 691 bp in length
*    22504    22603: gap of 100 bp
*    22604    23293: contig of 690 bp in length
*    23294    23393: gap of 100 bp
*    23394    24086: contig of 693 bp in length
*    24087    24186: gap of 100 bp
*    24187    24898: contig of 712 bp in length
*    24899    24998: gap of 100 bp
*    24999    25713: contig of 715 bp in length
*    25714    25813: gap of 100 bp
*    25814    26515: contig of 702 bp in length
*    26516    26615: gap of 100 bp
*    26616    27328: contig of 713 bp in length
*    27329    27428: gap of 100 bp
*    27429    28140: contig of 712 bp in length
*    28141    28240: gap of 100 bp
*    28241    28930: contig of 690 bp in length
*    28931    29030: gap of 100 bp
*    29031    29726: contig of 696 bp in length
*    29727    29826: gap of 100 bp
*    29827    30529: contig of 703 bp in length
*    30530    30629: gap of 100 bp
*    30630    31350: contig of 721 bp in length
*    31351    31450: gap of 100 bp
*    31451    32160: contig of 710 bp in length
*    32161    32260: gap of 100 bp
*    32261    32973: contig of 713 bp in length
*    32974    33073: gap of 100 bp
*    33074    33747: contig of 674 bp in length
*    33748    33847: gap of 100 bp
*    33848    34535: contig of 688 bp in length
*    34536    34635: gap of 100 bp
*    34636    35344: contig of 709 bp in length
*    35345    35444: gap of 100 bp
*    35445    36151: contig of 707 bp in length
*    36152    36251: gap of 100 bp
*    36252    36977: contig of 726 bp in length
*    36978    37077: gap of 100 bp
*    37078    37776: contig of 699 bp in length
*    37777    37876: gap of 100 bp
*    37877    38564: contig of 688 bp in length
*    38565    38664: gap of 100 bp
*    38665    39344: contig of 680 bp in length
*    39345    39444: gap of 100 bp
*    39445    40162: contig of 718 bp in length
*    40163    40262: gap of 100 bp
*    40263    40955: contig of 693 bp in length
*    40956    41055: gap of 100 bp
*    41056    41747: contig of 692 bp in length
*    41748    41847: gap of 100 bp
*    41848    42584: contig of 737 bp in length
*    42585    42684: gap of 100 bp
*    42685    43395: contig of 711 bp in length
*    43396    43495: gap of 100 bp
*    43496    44206: contig of 711 bp in length
*    44207    44306: gap of 100 bp
*    44307    45021: contig of 715 bp in length
*    45022    45121: gap of 100 bp
*    45122    45825: contig of 704 bp in length
*    45826    45925: gap of 100 bp
*    45926    46618: contig of 693 bp in length
*    46619    46718: gap of 100 bp
*    46719    47450: contig of 732 bp in length
*    47451    47550: gap of 100 bp

  Query Match           46.9%;  Score 58.6;  DB 2;  Length 53121;
  Best Local Similarity  67.2%;  Pred. No. 2.6;
  Matches   82;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  ||||  ||||| ||| | ||||| ||||| |||||  ||  | ||| |
Db      39920 CCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCC 39979

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
              |||| ||||||||||  |||| | |   ||| |||      || |  | ||   | | ||
Db      39980 GCCCCCCCCCGCGCCGCCCCCGCCCCCCGCCGCCCCCCGCGCGGCCGGCGCCCCCCGCGC 40039

Qy        121 CC 122
              ||
Db      40040 CC 40041
```

```
RESULT 10
AC116109
LOCUS       AC116109               68330 bp    DNA     linear   HTG 25-MAR-2002
DEFINITION  Mus musculus clone RP23-467G14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC116109
VERSION     AC116109.1  GI:19703220
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 68330)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-467G14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68330)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22810
                Center clone name: 467_G_14
            --------------.
            * NOTE: This record contains 85 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      647: contig of 647 bp in length
            *      648      747: gap of 100 bp
            *      748     1422: contig of 675 bp in length
            *     1423     1522: gap of 100 bp
            *     1523     2229: contig of 707 bp in length
            *     2230     2329: gap of 100 bp
            *     2330     3001: contig of 672 bp in length
            *     3002     3101: gap of 100 bp
            *     3102     3783: contig of 682 bp in length
            *     3784     3883: gap of 100 bp
            *     3884     4586: contig of 703 bp in length
            *     4587     4686: gap of 100 bp
            *     4687     5381: contig of 695 bp in length
            *     5382     5481: gap of 100 bp
            *     5482     6190: contig of 709 bp in length
            *     6191     6290: gap of 100 bp
            *     6291     7003: contig of 713 bp in length
            *     7004     7103: gap of 100 bp
            *     7104     7824: contig of 721 bp in length
            *     7825     7924: gap of 100 bp
            *     7925     8662: contig of 738 bp in length
            *     8663     8762: gap of 100 bp
            *     8763     9402: contig of 640 bp in length
            *     9403     9502: gap of 100 bp
            *     9503    10235: contig of 733 bp in length
            *    10236    10335: gap of 100 bp
            *    10336    11055: contig of 720 bp in length
            *    11056    11155: gap of 100 bp
            *    11156    11861: contig of 706 bp in length
            *    11862    11961: gap of 100 bp
            *    11962    12653: contig of 692 bp in length
            *    12654    12753: gap of 100 bp
            *    12754    13445: contig of 692 bp in length
            *    13446    13545: gap of 100 bp
            *    13546    14235: contig of 690 bp in length
            *    14236    14335: gap of 100 bp
            *    14336    15046: contig of 711 bp in length
            *    15047    15146: gap of 100 bp
            *    15147    15853: contig of 707 bp in length
            *    15854    15953: gap of 100 bp
            *    15954    16632: contig of 679 bp in length
            *    16633    16732: gap of 100 bp
            *    16733    17441: contig of 709 bp in length
            *    17442    17541: gap of 100 bp
            *    17542    18270: contig of 729 bp in length
            *    18271    18370: gap of 100 bp
            *    18371    19048: contig of 678 bp in length
            *    19049    19148: gap of 100 bp
            *    19149    19851: contig of 703 bp in length
            *    19852    19951: gap of 100 bp
            *    19952    20661: contig of 710 bp in length
            *    20662    20761: gap of 100 bp
            *    20762    21473: contig of 712 bp in length
            *    21474    21573: gap of 100 bp
            *    21574    22286: contig of 713 bp in length
            *    22287    22386: gap of 100 bp
            *    22387    23086: contig of 700 bp in length
            *    23087    23186: gap of 100 bp
            *    23187    23920: contig of 734 bp in length
            *    23921    24020: gap of 100 bp
            *    24021    24713: contig of 693 bp in length
            *    24714    24813: gap of 100 bp
            *    24814    25540: contig of 727 bp in length
            *    25541    25640: gap of 100 bp
            *    25641    26364: contig of 724 bp in length
            *    26365    26464: gap of 100 bp
            *    26465    27180: contig of 716 bp in length
            *    27181    27280: gap of 100 bp
            *    27281    27977: contig of 697 bp in length
            *    27978    28077: gap of 100 bp
            *    28078    28769: contig of 692 bp in length
            *    28770    28869: gap of 100 bp
            *    28870    29559: contig of 690 bp in length
            *    29560    29659: gap of 100 bp
            *    29660    30357: contig of 698 bp in length
            *    30358    30457: gap of 100 bp
            *    30458    31165: contig of 708 bp in length
            *    31166    31265: gap of 100 bp
            *    31266    31948: contig of 683 bp in length
            *    31949    32048: gap of 100 bp
            *    32049    32768: contig of 720 bp in length
```

```
            *    32769    32868: gap of 100 bp
            *    32869    33548: contig of 680 bp in length
            *    33549    33648: gap of 100 bp
            *    33649    34357: contig of 709 bp in length
            *    34358    34457: gap of 100 bp
            *    34458    35187: contig of 730 bp in length
            *    35188    35287: gap of 100 bp
            *    35288    35990: contig of 703 bp in length
            *    35991    36090: gap of 100 bp
            *    36091    36767: contig of 677 bp in length
            *    36768    36867: gap of 100 bp
            *    36868    37575: contig of 708 bp in length
            *    37576    37675: gap of 100 bp
            *    37676    38379: contig of 704 bp in length
            *    38380    38479: gap of 100 bp
            *    38480    39200: contig of 721 bp in length
            *    39201    39300: gap of 100 bp
            *    39301    40023: contig of 723 bp in length
            *    40024    40123: gap of 100 bp
            *    40124    40827: contig of 704 bp in length
            *    40828    40927: gap of 100 bp
            *    40928    41659: contig of 732 bp in length
            *    41660    41759: gap of 100 bp
            *    41760    42482: contig of 723 bp in length
            *    42483    42582: gap of 100 bp
            *    42583    43307: contig of 725 bp in length
            *    43308    43407: gap of 100 bp
            *    43408    44124: contig of 717 bp in length
            *    44125    44224: gap of 100 bp
            *    44225    44923: contig of 699 bp in length
            *    44924    45023: gap of 100 bp
            *    45024    45713: contig of 690 bp in length
            *    45714    45813: gap of 100 bp
            *    45814    46506: contig of 693 bp in length
            *    46507    46606: gap of 100 bp
            *    46607    47291: contig of 685 bp in length
            *    47292    47391: gap of 100 bp
            *    47392    48100: contig of 709 bp in length
            *    48101    48200: gap of 100 bp
            *    48201    48910: contig of 710 bp in length
            *    48911    49010: gap of 100 bp
            *    49011    49727: contig of 717 bp in length
            *    49728    49827: gap of 100 bp
            *    49828    50558: contig of 731 bp in length
            *    50559    50658: gap of 100 bp
            *    50659    51335: contig of 677 bp in length
            *    51336    51435: gap of 100 bp
            *    51436    52168: contig of 733 bp in length
            *    52169    52268: gap of 100 bp
            *    52269    52992: contig of 724 bp in length
            *    52993    53092: gap of 100 bp
            *    53093    53810: contig of 718 bp in length
            *    53811    53910: gap of 100 bp
            *    53911    54559: contig of 649 bp in length
            *    54560    54659: gap of 100 bp
            *    54660    55361: contig of 702 bp in length
            *    55362    55461: gap of 100 bp

  Query Match           46.9%;  Score 58.6;  DB 2;  Length 68330;
  Best Local Similarity  67.2%;  Pred. No. 2.4;
  Matches   82;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              |||| ||||  |||| ||||| ||||| ||||| ||||| |||||  ||  | ||| |
Db        352 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 411

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGCGGGCAGGCGGGGC 120
              ||| | ||| | ||   ||| | |  ||| ||| |  |||  |||||   |||||||
Db        412 CCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCGGGGCGGGGGCGCGGGGC 471

Qy        121 CC 122
              ||
Db        472 GC 473
```

```
RESULT 11
AC112672
LOCUS       AC112672               72645 bp    DNA     linear   HTG 09-MAY-2002
DEFINITION  Mus musculus clone RP24-418P12, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC112672
VERSION     AC112672.2  GI:20514911
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 72645)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-418P12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 72645)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 72645)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 9, 2002 this sequence version replaced gi:18860694.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
```

```
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L21384
    Center clone name: 418_P_12
--------------.
* NOTE: This record contains 89 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      700: contig of 700 bp in length
*      701      800: gap of 100 bp
*      801     1522: contig of 722 bp in length
*     1523     1622: gap of 100 bp
*     1623     2313: contig of 691 bp in length
*     2314     2413: gap of 100 bp
*     2414     3130: contig of 717 bp in length
*     3131     3230: gap of 100 bp
*     3231     3925: contig of 695 bp in length
*     3926     4025: gap of 100 bp
*     4026     4741: contig of 716 bp in length
*     4742     4841: gap of 100 bp
*     4842     5572: contig of 731 bp in length
*     5573     5672: gap of 100 bp
*     5673     6399: contig of 727 bp in length
*     6400     6499: gap of 100 bp
*     6500     7223: contig of 724 bp in length
*     7224     7323: gap of 100 bp
*     7324     8057: contig of 734 bp in length
*     8058     8157: gap of 100 bp
*     8158     8880: contig of 723 bp in length
*     8881     8980: gap of 100 bp
*     8981     9700: contig of 720 bp in length
*     9701     9800: gap of 100 bp
*     9801    10517: contig of 717 bp in length
*    10518    10617: gap of 100 bp
*    10618    11327: contig of 710 bp in length
*    11328    11427: gap of 100 bp
*    11428    12143: contig of 716 bp in length
*    12144    12243: gap of 100 bp
*    12244    12957: contig of 714 bp in length
*    12958    13057: gap of 100 bp
*    13058    13792: contig of 735 bp in length
*    13793    13892: gap of 100 bp
*    13893    14625: contig of 733 bp in length
*    14626    14725: gap of 100 bp
*    14726    15455: contig of 730 bp in length
*    15456    15555: gap of 100 bp
*    15556    16278: contig of 723 bp in length
*    16279    16378: gap of 100 bp
*    16379    17102: contig of 724 bp in length
*    17103    17202: gap of 100 bp
*    17203    17906: contig of 704 bp in length
*    17907    18006: gap of 100 bp
*    18007    18718: contig of 712 bp in length
*    18719    18818: gap of 100 bp
*    18819    19523: contig of 705 bp in length
*    19524    19623: gap of 100 bp
*    19624    20340: contig of 717 bp in length
*    20341    20440: gap of 100 bp
*    20441    21153: contig of 713 bp in length
*    21154    21253: gap of 100 bp
*    21254    21980: contig of 727 bp in length
*    21981    22080: gap of 100 bp
*    22081    22802: contig of 722 bp in length
*    22803    22902: gap of 100 bp
*    22903    23632: contig of 730 bp in length
*    23633    23732: gap of 100 bp
*    23733    24453: contig of 721 bp in length
*    24454    24553: gap of 100 bp
*    24554    25273: contig of 720 bp in length
*    25274    25373: gap of 100 bp
*    25374    26100: contig of 727 bp in length
*    26101    26200: gap of 100 bp
*    26201    26872: contig of 672 bp in length
*    26873    26972: gap of 100 bp
*    26973    27685: contig of 713 bp in length
*    27686    27785: gap of 100 bp
*    27786    28476: contig of 691 bp in length
*    28477    28576: gap of 100 bp
*    28577    29291: contig of 715 bp in length
*    29292    29391: gap of 100 bp
*    29392    30126: contig of 735 bp in length
*    30127    30226: gap of 100 bp
*    30227    30944: contig of 718 bp in length
*    30945    31044: gap of 100 bp
*    31045    31736: contig of 692 bp in length
*    31737    31836: gap of 100 bp
*    31837    32559: contig of 723 bp in length
*    32560    32659: gap of 100 bp
*    32660    33385: contig of 726 bp in length
*    33386    33485: gap of 100 bp
*    33486    34198: contig of 713 bp in length
*    34199    34298: gap of 100 bp
*    34299    35007: contig of 709 bp in length
*    35008    35107: gap of 100 bp
*    35108    35810: contig of 703 bp in length
*    35811    35910: gap of 100 bp
*    35911    36627: contig of 717 bp in length
*    36628    36727: gap of 100 bp
*    36728    37438: contig of 711 bp in length
*    37439    37538: gap of 100 bp
*    37539    38263: contig of 725 bp in length
*    38264    38363: gap of 100 bp
*    38364    39081: contig of 718 bp in length
*    39082    39181: gap of 100 bp
*    39182    39898: contig of 717 bp in length
*    39899    39998: gap of 100 bp
*    39999    40727: contig of 729 bp in length
*    40728    40827: gap of 100 bp
*    40828    41547: contig of 720 bp in length
*    41548    41647: gap of 100 bp
*    41648    42370: contig of 723 bp in length
*    42371    42470: gap of 100 bp
*    42471    43190: contig of 720 bp in length
*    43191    43290: gap of 100 bp
*    43291    43989: contig of 699 bp in length
*    43990    44089: gap of 100 bp
*    44090    44805: contig of 716 bp in length
*    44806    44905: gap of 100 bp
*    44906    45626: contig of 721 bp in length

  Query Match             46.6%;  Score 58.2;  DB 2;  Length 72645;
  Best Local Similarity   68.1%;  Pred. No. 2.6;
  Matches   81;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          7 CCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCGCCCGC 66
              ||||  |||| ||| || ||||| ||| |  |||| ||||| ||| |||||| | | | |
Db      52806 CCCCGCCCCCCGCCGCCCCCCCCCCCCGCCGCCCCGCCCCCGGCCCGCGCCCCCCCGCCC 52865

Qy         67 CCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGCCCAGC 125
              |||||||||   |||  | || |||  |||||  || |||  |       ||| ||| ||
Db      52866 CCCCGCGCCCCCCCCCCCCCGCCCCCGCCCGGCCCCCCCGGCGCGCCCGGGGCCCCGGC 52924

RESULT 12
```

```
AC020683
LOCUS       AC020683               80542 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-30L16, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC020683
VERSION     AC020683.2  GI:9140295
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 80542)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-30L16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 80542)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6682660.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4895
                Center clone name: 30_L_16
            --------------.
            * NOTE: This record contains 93 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      751: contig of 751 bp in length
            *      752      851: gap of 100 bp
            *      852     1603: contig of 752 bp in length
            *     1604     1703: gap of 100 bp
            *     1704     2545: contig of 842 bp in length
            *     2546     2645: gap of 100 bp
            *     2646     3412: contig of 767 bp in length
            *     3413     3512: gap of 100 bp
            *     3513     4285: contig of 773 bp in length
            *     4286     4385: gap of 100 bp
            *     4386     5154: contig of 769 bp in length
            *     5155     5254: gap of 100 bp
            *     5255     6018: contig of 764 bp in length
            *     6019     6118: gap of 100 bp
            *     6119     6884: contig of 766 bp in length
            *     6885     6984: gap of 100 bp
            *     6985     7755: contig of 771 bp in length
            *     7756     7855: gap of 100 bp
            *     7856     8625: contig of 770 bp in length
            *     8626     8725: gap of 100 bp
            *     8726     9469: contig of 744 bp in length
            *     9470     9569: gap of 100 bp
            *     9570    10338: contig of 769 bp in length
            *    10339    10438: gap of 100 bp
            *    10439    11225: contig of 787 bp in length
            *    11226    11325: gap of 100 bp
            *    11326    12098: contig of 773 bp in length
            *    12099    12198: gap of 100 bp
            *    12199    12973: contig of 775 bp in length
            *    12974    13073: gap of 100 bp
            *    13074    13831: contig of 758 bp in length
            *    13832    13931: gap of 100 bp
            *    13932    14698: contig of 767 bp in length
            *    14699    14798: gap of 100 bp
            *    14799    15569: contig of 771 bp in length
            *    15570    15669: gap of 100 bp
            *    15670    16447: contig of 778 bp in length
            *    16448    16547: gap of 100 bp
            *    16548    17303: contig of 756 bp in length
            *    17304    17403: gap of 100 bp
            *    17404    18172: contig of 769 bp in length
            *    18173    18272: gap of 100 bp
            *    18273    19043: contig of 771 bp in length
            *    19044    19143: gap of 100 bp
            *    19144    19900: contig of 757 bp in length
            *    19901    20000: gap of 100 bp
            *    20001    20769: contig of 769 bp in length
            *    20770    20869: gap of 100 bp
            *    20870    21648: contig of 779 bp in length
            *    21649    21748: gap of 100 bp
            *    21749    22522: contig of 774 bp in length
            *    22523    22622: gap of 100 bp
            *    22623    23389: contig of 767 bp in length
            *    23390    23489: gap of 100 bp
            *    23490    24262: contig of 773 bp in length
            *    24263    24362: gap of 100 bp
            *    24363    25128: contig of 766 bp in length
            *    25129    25228: gap of 100 bp
            *    25229    25989: contig of 761 bp in length
            *    25990    26089: gap of 100 bp
            *    26090    26861: contig of 772 bp in length
            *    26862    26961: gap of 100 bp
            *    26962    27737: contig of 776 bp in length
            *    27738    27837: gap of 100 bp
            *    27838    28612: contig of 775 bp in length
            *    28613    28712: gap of 100 bp
            *    28713    29477: contig of 765 bp in length
            *    29478    29577: gap of 100 bp
            *    29578    30327: contig of 750 bp in length
            *    30328    30427: gap of 100 bp
            *    30428    31195: contig of 768 bp in length
            *    31196    31295: gap of 100 bp
            *    31296    32066: contig of 771 bp in length
            *    32067    32166: gap of 100 bp
            *    32167    32942: contig of 776 bp in length
            *    32943    33042: gap of 100 bp
            *    33043    33818: contig of 776 bp in length
            *    33819    33918: gap of 100 bp
            *    33919    34680: contig of 762 bp in length
            *    34681    34780: gap of 100 bp
            *    34781    35554: contig of 774 bp in length
            *    35555    35654: gap of 100 bp
            *    35655    36412: contig of 758 bp in length
            *    36413    36512: gap of 100 bp
            *    36513    37292: contig of 780 bp in length
            *    37293    37392: gap of 100 bp
            *    37393    38153: contig of 761 bp in length
```

```
            *    38154    38253: gap of 100 bp
            *    38254    39027: contig of 774 bp in length
            *    39028    39127: gap of 100 bp
            *    39128    39895: contig of 768 bp in length
            *    39896    39995: gap of 100 bp
            *    39996    40756: contig of 761 bp in length
            *    40757    40856: gap of 100 bp
            *    40857    41618: contig of 762 bp in length
            *    41619    41718: gap of 100 bp
            *    41719    42502: contig of 784 bp in length
            *    42503    42602: gap of 100 bp
            *    42603    43392: contig of 790 bp in length
            *    43393    43492: gap of 100 bp
            *    43493    44191: contig of 699 bp in length
            *    44192    44291: gap of 100 bp
            *    44292    45068: contig of 777 bp in length
            *    45069    45168: gap of 100 bp
            *    45169    45934: contig of 766 bp in length
            *    45935    46034: gap of 100 bp
            *    46035    46808: contig of 774 bp in length
            *    46809    46908: gap of 100 bp
            *    46909    47676: contig of 768 bp in length
            *    47677    47776: gap of 100 bp
            *    47777    48527: contig of 751 bp in length
            *    48528    48627: gap of 100 bp
            *    48628    49315: contig of 688 bp in length
            *    49316    49415: gap of 100 bp
            *    49416    50173: contig of 758 bp in length
            *    50174    50273: gap of 100 bp
            *    50274    51010: contig of 737 bp in length
            *    51011    51110: gap of 100 bp
            *    51111    51877: contig of 767 bp in length
            *    51878    51977: gap of 100 bp
            *    51978    52754: contig of 777 bp in length
            *    52755    52854: gap of 100 bp
            *    52855    53621: contig of 767 bp in length
            *    53622    53721: gap of 100 bp
            *    53722    54493: contig of 772 bp in length
            *    54494    54593: gap of 100 bp
            *    54594    55361: contig of 768 bp in length
            *    55362    55461: gap of 100 bp
            *    55462    56221: contig of 760 bp in length
            *    56222    56321: gap of 100 bp
            *    56322    57083: contig of 762 bp in length
            *    57084    57183: gap of 100 bp
            *    57184    57966: contig of 783 bp in length
            *    57967    58066: gap of 100 bp
            *    58067    58814: contig of 748 bp in length
            *    58815    58914: gap of 100 bp
            *    58915    59691: contig of 777 bp in length
            *    59692    59791: gap of 100 bp
            *    59792    60561: contig of 770 bp in length
            *    60562    60661: gap of 100 bp
            *    60662    61416: contig of 755 bp in length
            *    61417    61516: gap of 100 bp

  Query Match             46.6%;  Score 58.2;  DB 2;  Length 80542;
  Best Local Similarity   65.0%;  Pred. No. 2.5;
  Matches   78;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          3 CCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGCGC 62
              ||  ||||  | ||  | ||| |||    |||| ||||| ||||| |||  | ||| | |
Db       2311 CCCCCCCCCCCGCCCNCGCCCCCCCNNNNCCCCCCCCCCCNCCCCCNGCCCCCTCCCCCCC 2370

Qy         63 CCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGCCC 122
              || ||||| | ||| ||  ||| | | ||  | |||  || |||||| | |  ||  |||
Db       2371 CCCCCCCCCCCCCGCGCGTGGCGCCGGCCNCCGCGGCNCCCCCGCGGCCCGCGGGCCCCC 2430


RESULT 13
AC016122
LOCUS       AC016122               84701 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-28B3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016122
VERSION     AC016122.2  GI:9136616
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 84701)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-28B3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 84701)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6456192.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4737
                Center clone name: 28_B_3
            --------------.
            * NOTE: This record contains 84 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      912: contig of 912 bp in length
            *      913     1012: gap of 100 bp
            *     1013     1911: contig of 899 bp in length
            *     1912     2011: gap of 100 bp
            *     2012     2917: contig of 906 bp in length
            *     2918     3017: gap of 100 bp
            *     3018     3903: contig of 886 bp in length
            *     3904     4003: gap of 100 bp
            *     4004     4913: contig of 910 bp in length
            *     4914     5013: gap of 100 bp
            *     5014     5919: contig of 906 bp in length
            *     5920     6019: gap of 100 bp
            *     6020     6926: contig of 907 bp in length
            *     6927     7026: gap of 100 bp
            *     7027     7926: contig of 900 bp in length
            *     7927     8026: gap of 100 bp
            *     8027     8961: contig of 935 bp in length
            *     8962     9061: gap of 100 bp
            *     9062     9973: contig of 912 bp in length
            *     9974    10073: gap of 100 bp
```

```
*    10074    10919: contig of 846 bp in length
*    10920    11019: gap of 100 bp
*    11020    11905: contig of 886 bp in length
*    11906    12005: gap of 100 bp
*    12006    12896: contig of 891 bp in length
*    12897    12996: gap of 100 bp
*    12997    13918: contig of 922 bp in length
*    13919    14018: gap of 100 bp
*    14019    14927: contig of 909 bp in length
*    14928    15027: gap of 100 bp
*    15028    15918: contig of 891 bp in length
*    15919    16018: gap of 100 bp
*    16019    16946: contig of 928 bp in length
*    16947    17046: gap of 100 bp
*    17047    17977: contig of 931 bp in length
*    17978    18077: gap of 100 bp
*    18078    19002: contig of 925 bp in length
*    19003    19102: gap of 100 bp
*    19103    19980: contig of 878 bp in length
*    19981    20080: gap of 100 bp
*    20081    20978: contig of 898 bp in length
*    20979    21078: gap of 100 bp
*    21079    22012: contig of 934 bp in length
*    22013    22112: gap of 100 bp
*    22113    23043: contig of 931 bp in length
*    23044    23143: gap of 100 bp
*    23144    24044: contig of 901 bp in length
*    24045    24144: gap of 100 bp
*    24145    25030: contig of 886 bp in length
*    25031    25130: gap of 100 bp
*    25131    26042: contig of 912 bp in length
*    26043    26142: gap of 100 bp
*    26143    27050: contig of 908 bp in length
*    27051    27150: gap of 100 bp
*    27151    28086: contig of 936 bp in length
*    28087    28186: gap of 100 bp
*    28187    29108: contig of 922 bp in length
*    29109    29208: gap of 100 bp
*    29209    30121: contig of 913 bp in length
*    30122    30221: gap of 100 bp
*    30222    31110: contig of 889 bp in length
*    31111    31210: gap of 100 bp
*    31211    32127: contig of 917 bp in length
*    32128    32227: gap of 100 bp
*    32228    33146: contig of 919 bp in length
*    33147    33246: gap of 100 bp
*    33247    34144: contig of 898 bp in length
*    34145    34244: gap of 100 bp
*    34245    35142: contig of 898 bp in length
*    35143    35242: gap of 100 bp
*    35243    36137: contig of 895 bp in length
*    36138    36237: gap of 100 bp
*    36238    37141: contig of 904 bp in length
*    37142    37241: gap of 100 bp
*    37242    38175: contig of 934 bp in length
*    38176    38275: gap of 100 bp
*    38276    39193: contig of 918 bp in length
*    39194    39293: gap of 100 bp
*    39294    40219: contig of 926 bp in length
*    40220    40319: gap of 100 bp
*    40320    41253: contig of 934 bp in length
*    41254    41353: gap of 100 bp
*    41354    42253: contig of 900 bp in length
*    42254    42353: gap of 100 bp
*    42354    43227: contig of 874 bp in length
*    43228    43327: gap of 100 bp
*    43328    44236: contig of 909 bp in length
*    44237    44336: gap of 100 bp
*    44337    45243: contig of 907 bp in length
*    45244    45343: gap of 100 bp
*    45344    46245: contig of 902 bp in length
*    46246    46345: gap of 100 bp
*    46346    47268: contig of 923 bp in length
*    47269    47368: gap of 100 bp
*    47369    48290: contig of 922 bp in length
*    48291    48390: gap of 100 bp
*    48391    49287: contig of 897 bp in length
*    49288    49387: gap of 100 bp
*    49388    50294: contig of 907 bp in length
*    50295    50394: gap of 100 bp
*    50395    51306: contig of 912 bp in length
*    51307    51406: gap of 100 bp
*    51407    52306: contig of 900 bp in length
*    52307    52406: gap of 100 bp
*    52407    53315: contig of 909 bp in length
*    53316    53415: gap of 100 bp
*    53416    54353: contig of 938 bp in length
*    54354    54453: gap of 100 bp
*    54454    55371: contig of 918 bp in length
*    55372    55471: gap of 100 bp
*    55472    56382: contig of 911 bp in length
*    56383    56482: gap of 100 bp
*    56483    57401: contig of 919 bp in length
*    57402    57501: gap of 100 bp
*    57502    58433: contig of 932 bp in length
*    58434    58533: gap of 100 bp
*    58534    59434: contig of 901 bp in length
*    59435    59534: gap of 100 bp
*    59535    60448: contig of 914 bp in length
*    60449    60548: gap of 100 bp
*    60549    61443: contig of 895 bp in length
*    61444    61543: gap of 100 bp
*    61544    62488: contig of 945 bp in length
*    62489    62588: gap of 100 bp
*    62589    63521: contig of 933 bp in length
*    63522    63621: gap of 100 bp
*    63622    64512: contig of 891 bp in length
*    64513    64612: gap of 100 bp
*    64613    65516: contig of 904 bp in length
*    65517    65616: gap of 100 bp
*    65617    66553: contig of 937 bp in length
*    66554    66653: gap of 100 bp
*    66654    67566: contig of 913 bp in length
*    67567    67666: gap of 100 bp
*    67667    68573: contig of 907 bp in length
*    68574    68673: gap of 100 bp
*    68674    69590: contig of 917 bp in length
*    69591    69690: gap of 100 bp
*    69691    70590: contig of 900 bp in length
*    70591    70690: gap of 100 bp
*    70691    71583: contig of 893 bp in length
*    71584    71683: gap of 100 bp
*    71684    72603: contig of 920 bp in length
*    72604    72703: gap of 100 bp
*    72704    73585: contig of 882 bp in length

  Query Match             46.4%;  Score 58;  DB 2;  Length 84701;
  Best Local Similarity   65.6%;  Pred. No. 2.6;
  Matches   82;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  | ||  ||||  | ||| ||||| ||||| ||||| |||||  ||  | |||||
Db      10657 CCCCCCCGCCCCCCCCGCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCGC 10716

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
               ||| ||||| | ||   ||| | | | | |||  | ||||| | |||  || | ||
Db      10717 CCCCCCCCCCCCCCCCTCCCCCCCNCCCCNCTTCCCCCTCCCGCCCCCGGCCCGCCGCGC 10776

Qy        121 CCAGC 125
                | ||
Db      10777 NCCGC 10781


RESULT 14
AX840138/c
LOCUS       AX840138               1416 bp    DNA     linear   PAT 16-DEC-2003
```

```
DEFINITION  Sequence 61 from Patent WO03078656.
ACCESSION   AX840138
VERSION     AX840138.1  GI:39978555
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Mosner,J. and Fries,H.R.
  TITLE     Dna markers for cattle or beef products
  JOURNAL   Patent: WO 03078656-A 61 25-SEP-2003;
            GAG Bioscience GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .1416
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
                     /note="Rind (SNP-Marker)"
ORIGIN

  Query Match             46.2%;  Score 57.8;  DB 6;  Length 1416;
  Best Local Similarity   66.4%;  Pred. No. 14;
  Matches   83;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  ||||  ||||| ||||| ||||| ||||| |||||  ||   | | |||
Db        448 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCGC 389

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
               |||  | ||||||  | |||||  | |  ||  ||||   ||||||  |   |   | |  |
Db        388 CCCCCGCGCCGCGCAGCGCCCGCCCCCCCCGCACCCCTCTCCGCCCCCCCCCCCCCCGCCC 329

Qy        121 CCAGC 125
              || ||
Db        328 CCCGC 324


RESULT 15
AX840139/c
LOCUS       AX840139               1416 bp    DNA     linear   PAT 16-DEC-2003
DEFINITION  Sequence 62 from Patent WO03078656.
ACCESSION   AX840139
VERSION     AX840139.1  GI:39978556
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Mosner,J. and Fries,H.R.
  TITLE     Dna markers for cattle or beef products
  JOURNAL   Patent: WO 03078656-A 62 25-SEP-2003;
            GAG Bioscience GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .1416
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
                     /note="Rind (SNP-Marker)"
ORIGIN

  Query Match             46.2%;  Score 57.8;  DB 6;  Length 1416;
  Best Local Similarity   66.4%;  Pred. No. 14;
  Matches   83;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy          1 CCCCTGCCCCTGCCCCTGCCCCCACCCCCACCCCCACCCCCACCCCCAGCCGGCGCCCGC 60
              ||||  ||||  ||||  ||||| ||||| ||||| ||||| |||||  ||   | | |||
Db        448 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCGC 389

Qy         61 GCCCGCCCCCGCGCCGGGCCCGGCTCGGCCCGACCCGGTTCCGCCGCGGGCAGGCGGGGC 120
               |||  | ||||||  | |||||  | |  ||  ||||   ||||||  |   |   | |  |
Db        388 CCCCCGCGCCGCGCAGCGCCCGCCCCCCCCCGCACCCCTCTCCGCCCCCCCCCCCCCCGCCC 329

Qy        121 CCAGC 125
              || ||
Db        328 CCCGC 324


Search completed: November  9, 2005, 02:15:38
Job time : 660.05 secs
```